THE ECONOMICS OF MUNICIPAL WATER RATES

By

ROBERT LEE GREENE

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA
1968

Dedicated to

Heather

ACKNOWLEDGEMENTS

It is impossible to acknowledge everyone who has contributed to this dissertation. However, there are several individuals to whom I must pay tribute.

I am deeply indebted to Dr. Milton Z. Kafoglis, my chairman, who gave freely of his time, patience, and knowledge in suggesting revisions in this work. I am particularly grateful to Dr. Kafoglis for his help in the development of several models including: the broader application of vertical summation, the methods of eliminating excess profits which lead to a quantity maximization conclusion, and the suburban tax problem.

Dr. Donald R. Escarraz devoted many of his evenings to helping me develop the interpretations of the economic models. These interpretations (particularly the Davidson and Hirshleifer interpretations) are my views, and not necessarily those of my chairman or my committee.

Appreciation is given to Dr. Ralph H. Blodgett and Dr. Clayton C. Curtis who, as members of my committee, gave me the encouragement needed for this undertaking.

I am indebted to the Institute of Government and the Department of Finance at the University of Georgia for making available the time necessary to complete this dissertation.

I also wish to acknowledge the loyalty of my wife Heather, and our two children, Rob and Tracy.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER	
I. INTRODUCTION	1
The Problem	1
Approach of the Study	4
Importance of the Study	9
Method and Outline of the Study	10
II. MARGINAL COST PRICING AND UTILITY RATE THEORY	14
Introduction	14
Development of the Marginal Cost Pricing Principle	15
Marginal Cost Pricing as a Policy	20
Limitations to Marginal Cost Pricing	24
Imperfect Competition	24
Externalities	25
Decreasing Costs	26
Interpersonal Comparisons and Equity	29
Joint Supply and Indivisibilities	33
An Alternative Solution to Jointness, Externality, Decreasing Costs, and Indivisibility	35
Interdependent Demands	39
Conclusions	40
III. THE THEORETICAL MODELS	43
Introduction	43
The Hotelling Model	44
The Steiner Model	46
The Hirshleifer Model	49
The Williamson Model	54
The Davidson Model	60
The Buchanan Model	67

TABLE OF CONTENTS (Continued)

The Underlying Assumptions 70

Joint Supply 71
Administrative Problems: Price Stability and Technological Deficiencies 74
Cost Functions 77

The Equity Implications 80

The Peak Problem 80
Cost Versus Ability to Pay 84
Elasticity of Demand and Resource Impact . . 84

Conclusions 87

IV. PRACTICAL DESIGN OF WATER RATES 89

Introduction 89

The Cost Allocation Technique 90

The Bias Towards Costs 90
The Patterson Allocation 95

Types of Charges 98

Fixed Charges 99
Variable Charges 103

An Evaluation of Current Practices 104

The Concepts of Equity and Efficiency 104
An Evaluation of the Cost Allocation Technique 107
Summary of Cost Allocation Evaluation 115
An Evaluation of Water Charges 116
The Lack of Zone Pricing 122

Conclusions 123

V. APPLIES MODELS 126

Introduction 126
Extension of the Model 128
The Economics of Block Pricing 133
Models which Interpret Present Practices . . 139
A Further Modification 144

TABLE OF CONTENTS (Continued)

Administrative Problems 146
A Fixed Charge Model 149
Zone Pricing 153
Conclusions 163

VI. DISTRIBUTIVE JUDGMENTS AND ECONOMIC EFFICIENCY 166

Introduction 166

Water Rates and Tax Policy 167

Tax Efficiency 168
Block Pricing and Taxation 173
Problems Associated with the Use of a "Water-Rate Tax" 178
Further Tax Efficiency Matters 180
The Suburban Tax Problem 182

Community Growth and Development Policy 186

Industrial Location 186
Rate Differentials and Community Development . 189

Conclusions 192

VII. CONCLUSIONS AND RECOMMENDATIONS 195

Conclusions 195
Recommendations 199

BIBLIOGRAPHY 204

LIST OF TABLES

TABLE Page

1. Zone Share of Capacity Costs 156

LIST OF FIGURES

Figure		Page
I.	A Firm with Decreasing Costs	27
II.	Vertical Summation	36
III.	The Hotelling Model	45
IV.	The Steiner Model	47
V.	Hirshleifer's Continuous Cost Model	51
VI.	Hirshleifer's Discontinuous Cost Formulation	53
VII.	The Williamson Model	56
VIII.	The Williamson Model Assuming Indivisibility	58
IX.	The Davidson Model	61
X.	An Interpretation of the Davidson Solution	62
XI.	A Total Cost Interpretation of the Davidson Model	63
XII.	The Average Cost Interpretation of the Davidson Model	65
XIII.	The Optimal Solution in the Absence of Time Jointness	72
XIV.	The Vertical Summation of Intra-Cycle Demands	73
XV.	Traditional Efficiency Pricing with Interdependent Demands	75
XVI.	Cost Functions with a Capacity Constraint	78
XVII.	An Hourly Load Chart for Three Users	81
XVIII.	An Hourly Load Chart	83
XIX.	Elasticity and the Impact of Price	85
XX.	The Combined Benefits Derived from Water Capacity	100

LIST OF FIGURES (Continued)

Figure		Page
XXI.	Long-Run Costs of Monopoly Firm Producing Water with Constant Factor Prices	106
XXII.	An Hourly Load Chart	109
XXIII.	Determination of the Peak and Off-Peak Demands . .	113
XXIV.	The Williamson Model	129
XXV.	Non-Discriminating Monopoly	130
XXVI.	First Degree Price Discrimination	133
XXVII.	A Block-Rate Model	136
XXVIII.	Williamson's Model with Block Pricing	137
XXIX.	A Monopoly under Different Pricing Alternatives .	141
XXX.	An Interpretation of Utility Pricing and Investment Practices	145
XXXI.	Fixed Charge Based upon the Ability to Congest the Water System	150
XXXII.	Capacity Costs and Use Characteristics of Three Zones	155
XXXIII.	The Marginal Solution for Serving Three Zones . .	161
XXXIV.	A Municipal Water Utility with a Profit Restraint.	169
XXXV.	A Per-Unit Tax versus an Ad Valorem Tax	171
XXXVI.	A Five-Block-Rate Schedule	175
XXXVII.	Supply and Demand for Public Goods	184

CHAPTER I

INTRODUCTION

The Problem

In the last decade, there has been growing concern at the policy level about the rate practices of municipally owned water utilities. To an increasing extent, this concern has been reflected in the applied literature and has centered on the relation of municipal water rate structures to the efficiency of resource use and to various criteria of equity or "reasonableness." Thus, writers have generated two basic questions: (1) do present municipal water rate practices contribute to economic efficiency?; and (2) are present water rate practices equitable?[1]

Economic efficiency is concerned with the attainment of that allocation of resources, or "input-output mix," which maximizes the satisfactions of the consumers in the economy. This efficiency criterion requires each water user to pay a price which reflects the marginal costs

---

[1]Jack Hirshleifer, James C. DeHaven, and J.W. Milliman, Water Supply (Chicago: University of Chicago Press, 1960), pp. 161-162; J.C. Bonbright, "Fully Distributed Costs in Utility Rate Making," American Economic Review, LI (May, 1961), p. 312; Irving K. Fox and Orric C. Herfindahl, "Attainment of Efficiency in Satisfying Demands for Water Resources," American Economic Review, LIV (May, 1964), p. 205; Gordon P. Fisher, "New Look at Resources Policy," Journal of the American Water Works Association, LVII (March, 1965), p. 359.

he imposes, assuming there are no complications stemming from joint supply, interdependent demands, externalities, and distributional objectives in conflict with economic efficiency.

While economic efficiency reflects the orientation of economists, water engineers seem to have a different set of criteria. Their major concern centers around the ability of the rate structure to recover total costs. Thus, water rates are established so that total revenue is at least equal to total costs. This criterion of total cost recovery is combined with a distributional criterion which requires each user to pay the "full" costs he imposes upon the utility.[2]

Utility engineers appear to pay little attention to the formal distinction between the efficiency and the distributional aspects of the rate structures they develop. It seems that water engineers equate a dubious concept of economic efficiency with distributional equity and try to solve both problems simultaneously. Thus, rate structures which yield total revenue equal to total cost are considered both equitable and efficient. Utility engineers do not appear to consider the

---

[2]For articles reflecting these different definitions and criteria see the following: William G. Shepherd, "Marginal Cost Pricing in American Utilities," Southern Economic Journal, XXIII (July, 1966), p. 60; Hirshleifer, et al., Water Supply, p. 162; Staff Report, "The Water Utility Industry in the United States," Journal of the American Water Works Association, LVIII (July, 1966), p. 772; E.D. Bonine, "Making a Water Utility Solvent," Journal of the American Water Works Association, XLV (May, 1953), p. 457; Bonbright, "Fully Distributed Costs in Utility Rate Making," pp. 305-12; J.C. Bonbright, "Two Partly Conflicting Standards of Reasonable Utility Rates," American Economic Review, XLVIII (May, 1957), pp. 386-93.

possibility that a utility might have to operate with profits or losses to satisfy the criteria of economic efficiency, nor do they seem to really appreciate the significance of marginal analysis.[3] In short, they employ a full cost allocation as a criterion of both efficiency and equity. This study examines the possibility of applying the criteria suggested by modern welfare economists in the hope that some measure of contribution might be made toward the development of improved water rate structures.

There is little in the present literature which evaluates the alternative criteria a municipally owned water utility might use in determining water rate practices. Consequently, there is no analysis of the economic implications of alternative criteria. A modest contribution to the field entails recognition of the various alternative criteria and an analysis of their impact upon economic efficiency. The welfare economist is in a position to provide such analysis, as ably stated by H. Thomas Koplin:

> It is true that equity and environment are subjective factors, that they rest on personal values, and the economist has no claim to superior values. But it is equally true that both are as important

---

[3]See: Louis R. Howson, "Review of Ratemaking Theories," Journal of the American Water Works Association, LVIII (July, 1966), p. 855; William L. Patterson, "Practical Water Rate Determination," Journal of the American Water Works Association, LIV (August, 1962), p. 906; Jerome W. Milliman, "The New Price Policies for Municipal Water Service," Journal of the American Water Works Association, LVI (February, 1964), p. 127.

> in determining policy, and human satisfactions, as is efficiency. It is therefore not only proper but essential that the economist incorporate them in his policy analysis. In doing so he will simply be catching up with regulators, not to mention the public.[4]

Welfare economics lends itself to this type of policy analysis.

Approach of the Study

The problem of maximizing welfare is one of combining economic efficiency with distributional equity in such a manner that the well-being of the individuals in the community is maximized. In the standard Paretian sense, economic efficiency is achieved when resources are allocated so that it is impossible to increase the welfare of one individual without decreasing the well-being of some other individual.[5] In other words, a situation is "Pareto inefficient" so long as it is possible to move to another situation and in the process make at least one person better off without making some other person worse off. Efficiency, thus defined, is achieved when (1) the consumer equates the exchange value of the last unit purchased with the production value of that unit, and (2) the exchange value of the last unit purchased is equal

---

[4] H. Thomas Koplin, "Discussion," American Economic Review, LI (May, 1961), p. 336.

[5] For the use of this criterion in the Paretian context see: William J. Baumol, Welfare Economics and the Theory of the State (Cambridge: Harvard University Press, 1965), pp. 163-79.

for all consumers.[6] These criteria are fulfilled when price is equal to marginal cost. However, for policy prescription, it is also necessary that the income distribution which results from marginal cost pricing be deemed desirable. If the resulting distribution of income is considered undesirable, then a departure from the marginal cost pricing policy may be required.[7] Any departure, however, entails distributional judgments which will have an impact upon the attainment of economic efficiency as the allocation fails to satisfy the two basic criteria.

---

[6]For the presentation of these criteria see: Abba P. Lerner, The Economics of Control (New York: The Macmillan Co., 1944), pp. 7-136; William J. Baumol, Economic Theory and Operations Analysis (2nd ed.; Englewood Cliffs: Prentice-Hall, 1965), pp. 355-63; Abram Bergson, Essays in Normative Economics (Cambridge: Harvard University Press, 1966), pp. 78-90; A. Reder, Studies in the Theory of Welfare Economics (New York: Columbia University Press, 1947), Chapter 2; J. Hirshleifer and J.W. Milliman, "Urban Water Supply: A Second Look," American Economic Review, LVII (May, 1967), pp. 169-78; Oliver E. Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," American Economic Review, LVI (September, 1966), p. 812; Harvey Averch and Leland L. Johnson, "Behavior of the Firm under Regulatory Constraint," American Economic Review, LII (December, 1962), p. 1052; F.P. Linaweaver and John C. Geyer, "Use of Peak Demands in Determination of Residential Rates," Journal of the American Water Works Association, LVI (April, 1964), p. 413; Hirshleifer, et al., Water Supply, p. 87.

[7]See the following works: Paul A. Samuelson, Foundations of Economic Analysis (New York: Atheneum, 1965), p. 253; Hirshleifer, et al., Water Supply, p. 90; Linaweaver and Geyer, "Use of Peak Demands in Determination of Residential Rates," p.417; Shepherd, "Marginal Cost Pricing in American Utilities," pp. 59-60.

Some economists, notably John R. Hicks, maintain that judgments concerning the distribution of income do not have to be made even for a policy prescription. Since any change from an inefficient to an efficient solution has the "potential" of making everyone better off, the distributional problem may, according to this view, be avoided.[8] However, "actual" welfare and "potential" welfare are different things, and Samuelson, Little, and others, maintain that the distributional problem cannot be avoided at the policy level.[9] The writer, at least for purposes of this applied study, holds to the view that a policy recommendation should consider both tests.

A municipal water utility is embroiled in decisions which entail distributional or equity judgments. The water utility, as part of the community's available revenue sources, is manipulated by local officials to create tax equity (or inequity), to acquire additional general revenue, and to carry out community growth and development programs. What these officials may fail to recognize is the conflict that sometimes arises between economic efficiency and distributional equity. Rate structures which satisfy the criteria for economic efficiency are not

---

[8]John R. Hicks, "The Foundations of Welfare Economics," Economic Journal, XLIX (December, 1939), pp. 696-712; Nicholas Kaldor, "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," Economic Journal, XLIX (September, 1939), pp. 549-52.

[9]Paul A. Samuelson, "Welfare Economics and International Trade," American Economic Review, XXVIII (June, 1938), pp. 261-66; Samuelson, Foundations of Economic Analysis, pp. 249-52; I.M.D. Little, A Critique of Welfare Economics (Oxford: Oxford University Press, 1958), pp. 217-38.

always the most equitable. On the other hand, rates considered to be equitable may not satisfy the criteria for economic efficiency. The problem facing the municipal water utility is one of combining economic efficiency with distributional equity in such a manner as to maximize the welfare of the people in the community.[10] The problem becomes one of finding an optimal size plant and an optimal rate structure without impeding, and hopefully encouraging, improvements in the distribution of income.

However, when economic efficiency and distributional equity conflict in the establishment of municipal water rate practices, a trade-off between the two becomes necessary. The existence of such conflict is evidenced by the presence of conflicting practices in present water rate structures which represent failures to resolve the conflict. For example, a utility may offer a discount for water used during the off-peak period in an attempt to reduce the peak load problem. At the same time, the rate structure will incorporate a low promotional rate which may encourage peak use. Obviously, the two are in conflict.

Efficiency and equity also come into conflict when a utility employs a declining block-rate structure with a rate below marginal cost in the last block. This practice is often used to promote industrial

---

[10] A more precise definition of welfare maximization is presented in Chapter II.

location.[11] Since low rates benefit large users, their effect is greater capacity costs for the utility. The utility's costs increase as the ratio of peak use to off-peak use increases. Therefore, a decision to sell at below marginal cost results in both economic inefficiency--expansion of capacity to sell a greater output at below marginal cost--and distributional effects, since the over expansion must be paid for by someone.

Marginal cost pricing, as a single pricing policy, however, is not always applicable to a municipally owned utility because the assumptions underlying the marginal cost principle are not always satisfied. The water utility is faced with a time jointness problem that arises because of peak and off-peak water use.[12] Attempts to resolve this problem are further complicated because peak and off-peak demands may be interdependent, particularly on an hourly basis. In addition, there are externalities, such as public health and sanitation, which must be taken into account.[13] When these conditions arise, a strict adherence to

---

[11] Low promotional rates are also used to generate greater water demand from present water users. These low rates lead to greater lawn sprinkling and uncontrolled air conditioning. The low rates, therefore, conflict with water scarcities arising during the summer months and the overall effect is to put greater demand upon the system capacity.

[12] For an adequate discussion of the problems of joint supply see: Donald H. Wallace, "Joint Supply and Overhead Costs and Railway Rate Policy," Quarterly Journal of Economics, XLVIII (August, 1934), pp. 583-619; A.C. Pigou, The Economics of Welfare (4th ed.; London: The Macmillan and Co., Ltd., 1932), pp. 297-99, 300-301.

[13] A classification of externalities is given by: Tibor Scitovsky, "Two Types of Externalities," Journal of Political Economy, LXII (April, 1954), pp. 143-51; Milton Z. Kafoglis, Welfare Economics and Subsidy Programs (Gainesville: University of Florida Press, 1961), pp. 16-38.

marginal cost pricing is not feasible or desirable. Indeed, these complications shall occupy the bulk of the attention of this study.

Importance of the Study

This study recognizes a need for additional analysis of municipal water rate practices in relation to economic criteria for determining plant capacity and rate structures.

The need for additional analysis is especially urgent in the light of the current development of congestion of most public facilities, and the problems faced by local governments such as the need for increased general revenues, tax equity, and community growth and development. Since these problems require solutions involving distributional judgments, traditional theory puts them aside, consequently failing to provide a complete analysis of some of the more important factors determining municipal water rate practices. Distributional decisions which are incorporated into the rate structures must be isolated and, if possible, evaluated if their impacts are to be revealed and taken into account in the development of policy.

In determining their water rate practices, municipally owned utilities do not seem to follow any consistent criteria. Distributional judgments are built into the rate structures in a haphazard, ad hoc, manner.[14] Some of the more common criteria by which rates are evaluated

---

[14]James M. Buchanan, "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXX (August, 1966), pp. 463-71.

are (1) average costs,[15] (2) rates which will maximize revenue with least resistence from members of the community,[16] (3) rate structures developed by neighboring communities,[17] and (4) rates which will generate industrial location. What is lacking is a series of broad pricing guidelines that municipally owned utilities can follow in the determination of rate structures and plant size which embody the effects of distributional decisions. The intent of this study is not to determine specific rate structures under alternative conditions. Rather, the intent is to arrive at a set of guidelines which will achieve efficiency, with equity considerations temporarily set aside. Rate practices based upon these guidelines will enable those who make rate decisions to realize the efficiency implications of their distributional judgments when equity and efficiency come into conflict. The need is to highlight and delineate the nature of the conflict in the hope that this might encourage superior rate policies.

Method and Outline of the Study

The hypothesis of this study is that additional economic analysis of water rate practices can provide a basis for policy changes which

---

[15]This criterion seems to be the most common. See: Hirshleifer, et al., Water Supply, p. 111; Howson, "Review of Ratemaking Theories," p. 850.

[16]Samuel S. Baxter, "Principles of Rate Making for Publically [sic] Owned Utilities," Journal of the American Water Works Association, LII (October, 1960), p. 1227.

[17]Raymond J. Faust, "The Needs of Water Utilities," Journal of the American Water Works Association, LI (June, 1959), p. 703.

will make rates not only more efficient, but also more equitable.

The study is an applied theoretical analysis which can be divided into four basic parts. The first part, which entails Chapter II, is an evaluation of marginal cost pricing as a basis for rate and capacity determination for municipally owned water utilities. This part contrasts the Paretian framework with the "old" welfare economics of Marshall, Pigou, Lerner,and others. This section also examines the assumptions underlying the marginal cost pricing principle--independent demands, the absence of externalities, the absence of joint supply, and the distributional aspect--to determine the extent to which it can serve as the basis for improved municipal water rate practices.

The second section, which includes Chapter III, is an evaluation of the theoretical models developed by recent writers including Buchanan, Davidson, Hirshleifer, Hotelling, Steiner, and Williamson, which have appeared in the current literature.[18] This evaluation seeks to determine the validity of these models as useful guidelines for municipal water rate practice. These models are found to be lacking and in need of further modification and development if they are to be useful at the applied

---

[18]Buchanan, "Peak Loads and Efficient Pricing: Comment," pp. 463-71; Jack Hirshleifer, "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXII (August, 1958), pp. 451-62; Ralph K. Davidson, Price Discrimination in Selling Gas and Electricity (Baltimore: Johns Hopkins University Press, 1955); Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica, VI (July, 1938), pp. 242-69; Peter O. Steiner, "Peak Loads and Efficient Pricing," Quarterly Journal of Economics, LXXI (November, 1957), pp. 585-610; Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," pp. 810-27.

level. Modifications are suggested by relaxing some of the underlying assumptions, and then examining and developing the implications of the new assumptions.

Chapter IV constitutes the third section which is an evaluation of water rate practice as it is used and advocated in the day-to-day operations of water utilities. A model developed by William L. Patterson is examined closely since it represents a typical approach to the allocation of costs advocated by water rate analysts.[19] This section probes the various cost classifications and rate classifications used by water utilities to discern their limitations as well as their efficiency and equity implications. It also critically examines the various types of charges employed by water utilities.

The fourth part, which includes Chapters V and VI, is a synthesis of present theory and practice, which leads to some policy recommendations concerning the type of water rate practices that might be used. The intent is to derive practical water rate guidelines using economic analysis under different sets of assumptions. Chapter V examines the welfare implications of alternative models such as simple and discriminating monopoly, block-rate schedules, quantity maximization and out-of-pocket cost pricing, and zone pricing. Chapter VI analyzes and evaluates the efficiency and equity implications of the various

---

[19]Patterson, "Practical Water Rate Determination," p. 909.

alternatives used in water rate structures to achieve such objectives as tax equity, greater general fund revenues, and community growth and development.

CHAPTER II

MARGINAL COST PRICING AND UTILITY RATE THEORY

Introduction

Modern public utility rate theory may be contrasted sharply with traditional rate practice. The essence of traditional rate practice has been the protection of consumers from monopoly exploitation through public regulation, and the emphasis has been placed on the attainment of reasonableness or equity in the relationship between buyer and seller. This emphasis also characterized the writings of early economists who emphasized the reasonableness of both rates and profits.[1] Thus, utility rate theory developed as a separate compartment of applied economic theory and revolved around the valuation of the rate base and unjust rate discrimination. In recent years, there has been a marked shift in the emphasis of both theory and practice from standards of equity and reasonableness to more or less objective standards of efficiency in the development of utility rate structures. This change has brought public utility rate theory into a much closer relationship with orthodox Marshallian price theory.

This chapter briefly summarizes the historical development and conceptual basis of marginal cost pricing. In addition, it organizes some of the complex situations--indivisibility, joint supply, externality--which create difficulties at both the theoretical and applied levels.

---

[1] Adam Smith, The Wealth of Nations, Modern Library (New York: Random House, Inc., 1937), pp. 681-716.

Although each of these complexities has been discussed separately by various writers in relation to specific problems, it is felt that a modest effort to pull the strands together will provide a more complete picture of the problems associated with marginal cost pricing and utility rate theory--both theoretical and applied--than exists at the present time.

Development of the Marginal Cost Pricing Principle

Alfred Marshall was the first economist to develop a sophisticated model for the analysis of economic efficiency. Marshall advocated the use of marginal analysis as a means of analyzing economic efficiency and suggested consumers' surplus as a normative criterion.[2] According to Marshall, the equality of supply and demand characterized efficiency in resource use. Moreover, this equality led to the maximization of consumers' surplus. Marshall's discussion of increasing and decreasing costs is a classic example of the use of marginal analysis in evaluating the welfare effects of alternative situations.[3] Employing the maximization

---

[2]Although the welfare significance of consumers' surplus is still debated, it seems to have better standing today among many writers than even Marshall attributed to it. See: David M. Winch, "Consumer's Surplus and the Compensation Principle," American Economic Review, LV (June, 1965), pp. 395-423; John R. Hicks, "The Rehabilitation of Consumers' Surplus," Review of Economic Studies, VIII (February, 1941), pp. 108-16.

[3]For Marshall's discussion see: Alfred Marshall, Principles of Economics (8th ed.; London: Macmillan and Co., Ltd., 1920), pp. 390-94.

of consumers' surplus as a criterion, Marshall analyzed the policy of taxing firms experiencing decreasing returns (increasing costs) and paying subsidies to firms experiencing increasing returns (decreasing costs). Marshall proposed that the output of firms operating under increasing costs be contracted because at a profit-maximizing industry equilibrium, marginal cost exceeds average cost. On the other hand, the output of firms operating under decreasing costs should be expanded.[4] Marshall's classic tax-bounty analysis is recognized as the original marginal cost theory of pricing.

The concept of consumers' surplus developed by Marshall implied an interpersonal comparison of utility. Marshall himself recognized that the "surplus," measured under a market demand function, measures satisfaction only if one passes over the possibility that a given sum of money gives different degrees of satisfaction to different people.[5]

Following in Marshall's footsteps, Pigou formalized the conditions necessary for the maximization of welfare.[6] Pigou, substituting social net product or the national dividend in place of consumer's surplus

---

[4]For a summary of this analysis see: Milton Z. Kafoglis, Welfare Economics and Subsidy Programs (Gainesville: University of Florida Press, 1961), pp. 9-10.

[5]This point is well developed by: Nancy Ruggles, "The Welfare Basis of the Marginal Cost Pricing Principle," Review of Economic Studies, XVII (1949-1950), p. 31.

[6]A.C. Pigou, The Economics of Welfare (4th ed.; London: Macmillan and Co., Ltd., 1932), p. 290.

as the measure of welfare, established a set of marginal conditions which must be satisfied to achieve a welfare maximum. Nancy Ruggles states:

> Pigou in effect implied a set of marginal conditions of production when he said that the marginal net social product of resources in each use must be equal in order to maximize the national dividend. His proof is the demonstration that if the marginal conditions are not met, the aggregate national dividend could be increased by removing resources from uses in which the marginal social net product is lower and employing them in uses with higher social net products.[7]

By virtue of using the national dividend as the measure of welfare, Pigou also built interpersonal utility comparisons into his theory because the value of the national dividend depends upon the distribution of income. That is, any single bundle of goods will be valued differently depending upon the distribution of income. Pigou recognized the problem and, accordingly, employed both the size and the distribution of the national dividend as coordinate measures of welfare which have to be traded off.[8] According to Pigou:

> On this basis, it is desired, if possible, to establish some connection between changes in the distribution of the national dividend and changes in economic welfare, corresponding to the connection established in the preceding chapter between changes in the size of the

---

[7] Ruggles, "The Welfare Basis of the Marginal Cost Pricing Principle," p. 32.

[8] Ibid., p. 32.

national dividend and changes in economic welfare.[9]

The problem of interpersonal comparisons was eventually formalized rigorously in the works of Pareto. Through the use of ordinalist assumptions and marginal analysis, Pareto developed a welfare criterion which avoided cardinal utility as well as the necessity of interpersonal comparisons of utility. Pareto defined the welfare maximum as an arrangement where it is not possible to make everyone better off by any movement within the system.[10]

Throughout the works of the early theorists, there was a gradual development of the concept of welfare along with the marginal conditions necessary to achieve the welfare maximum.[11] Marshall, Pigou, Lerner,

---

[9]Pigou, The Economics of Welfare, p. 89.

[10]Rugglès, "The Welfare Basis of the Marginal Cost Pricing Principle," p. 32.

[11]The seven marginal conditions are: (1) the marginal rates of substitution between two commodities should be equal for any two consumers; (2) the marginal rates of transformation between two commodities are equal for all producers; (3) the marginal physical product of a given factor for a given commodity is equal for all commodity producers; (4) the marginal rates of equal-product substitution of two factors are the same for all producers; (5) the marginal rate of indifferent substitution of any consumer for two products is equal to the marginal rate of transformation of these two commodities in production; (6) the marginal reward of a factor equals the marginal rate of substitution of reward for use; and (7) the marginal rates of time substitution of two individuals for a given asset are equal. This listing and the implications are given by: Kenneth E. Boulding, "Welfare Economics," A Survey of Contemporary Economics, ed. by Bernard F. Haley, I (Homewood: Richard D. Irwin, Inc., 1952), pp. 1-34. It should be noted that satisfaction of the seven marginal conditions does not imply that welfare is maximized. The conditions can define a minimum as well as a maximum. Moreover, the maximum they do define can only be a relative maximum and

and others, made use of interpersonal comparisons, while Pareto and his followers avoided such essentially subjective valuations. Both schools of thought developed identical marginal conditions, but there was, and still is, a serious difference of opinion as to their welfare significance.[12] Nevertheless, the marginal conditions which evolved are not a source of controversy and are a familiar part of the work of all welfare economists. Moreover, the development of these conditions has provided the basis for a distinction between economic efficiency and distributional equity. Basically, this distinction is the difference between the "old" and the "new" welfare economics. The choice between the two approaches does not involve a choice between equity and efficiency as the proper objective of economic analysis, but does involve a choice relating to the validity of the distinction between the two. By isolating the difficult distributional question, the "new," or Paretian, welfare economists have developed a single pricing principle--price equal to

---

not an absolute maximum. The conditions do, however, have to be satisfied to maximize total welfare. They are necessary, but not sufficient, conditions which must be combined with the so-called "total" conditions which remain poorly defined. See: John R. Hicks, Value and Capital (2nd ed.; Oxford: Oxford University Press, 1946), pp. 62-77; George J. Stigler, The Theory of Price (rev. ed.; New York: The Macmillan Co., 1962), pp. 42-95.

[12]For discussions of these points see: Ruggles, "The Welfare Basis of the Marginal Cost Pricing Principle," pp. 29-40; I.M.D. Little, A Critique of Welfare Economics (Oxford: Oxford University Press, 1958), pp. 67-216; William J. Baumol, Economic Theory and Operations Analysis (2nd ed.; Englewood Cliffs: Prentice-Hall, 1965), pp. 355-85.

marginal cost--as a policy guide. Obviously, this approach is very limited where difficult distributional problems exist and must, in most instances, be supplemented by some sort of distributional judgment.

Marginal Cost Pricing as a Policy

Pigou was the first to develop the utility rate theory implied in the marginal conditions, and he applied the criteria to the railroad industry.[13] Pigou concluded that price equal to marginal cost is a completely general guide to economic efficiency and went on to evaluate the implications of divergences between price and marginal cost.[14] In his analysis, Pigou struggled with the problem of applying marginal cost pricing to decreasing cost industries. In these instances, the difficulty of applying the principle is that prices equal to marginal cost will fail to recover the total costs of operation. Losses arise because average costs are higher than marginal costs which have been

---

[13]Many of the important points arose during the controversy between Pigou and Taussig in 1913 over the question of joint supply. See the following: A.C. Pigou, The Economics of Welfare, pp. 290-317; F.W. Taussig, "Railway Rates and Joint Costs Once More," Quarterly Journal of Economics, XXVII (February, 1913), pp. 378-84; A.C. Pigou, "Railway Rates and Joint Costs," Quarterly Journal of Economics, XXVII (May, 1913), pp. 535-36 and the rebuttal immediately following by F.W. Taussig, Quarterly Journal of Economics, XXVII (May, 1913), pp. 536-38; A.C. Pigou, "Railway Rates and Joint Costs," Quarterly Journal of Economics, XXVII (August, 1913), pp. 687-92 and the rebuttal immediately following by F.W. Taussig, Quarterly Journal of Economics, XXVII (August, 1913), pp. 692-93.

[14]Pigou, The Economics of Welfare, pp. 381-408.

equated with price. If the firm does not receive some type of bounty, or negative tax, services cannot be provided in the long-run. In this case, if the community wants the commodity to be produced in optimal quantities, it will have to subsidize the producer, or take over the operation of the firm.

Much of the early work on marginal cost pricing revolved around the problems of welfare concepts, joint costs, decreasing costs, and distributional equity. The early works developed terminology such as Marshall's surplus, Pigou's social net product, and the marginal equalities which characterized a welfare maximum.

In 1938, the appearance of Harold Hotelling's article "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates" marked a major breakthrough in the use of marginal cost pricing as a policy prescription to be applied to public utilities.[15] Along with Hotelling's article, a controversy arose concerning the use of marginal cost pricing as a practical basis for public utility regulation. In a broader sense, the "marginal cost controversy," which arose during the late 1930's, represented the response of neo-classical welfare economics to the problems of unemployment, excess capacity, and income distribution that developed during the Great Depression. Although the controversy exposed and illuminated many theoretical and applied issues,

---

[15]Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica, VI (July, 1938), pp. 242-69.

the center of the stage, during the period 1940-1960, was held by Keynesian economics; the problems of efficient pricing remained in the background. The publication of Arrow's impossibility theorem in 1949 further weakened the theoretical case for marginal cost pricing.[16] Rehabilitation began in 1952 with the publication of Baumol's Welfare Economics and the Theory of the State, which developed the concept of externality in a Paretian framework. Recent contributions by Oliver E. Williamson, James M. Buchanan, and others, have added impetus to the development of an applied welfare economics which combines the concept of externality with the marginal cost criterion to provide a reasonably rigorous framework for policy prescriptions. It is important to note Hotelling's significant conclusions, and the controversy which arose after the article appeared.

Hotelling set forth and rigorously defended the proposition that marginal cost pricing is the proper policy to follow in the determination of utility rates.[17] Marginal cost pricing was now set forth as a definite policy rigorously derived from highly formal welfare conditions. Like Pigou, Hotelling discussed the problem of applying marginal cost pricing to firms operating with decreasing costs. Hotelling's position was that such firms should receive compensation from the

---

[16]Kenneth J. Arrow, Social Choice and Individual Values (New York: John Wiley and Sons, 1951), p. 26.

[17]Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," p. 242.

community. He recommended that the compensation (subsidy) be raised through a general income tax.[18]

The tax criteria developed by Hotelling demonstrated that, for a community to maximize welfare, it was necessary to tax income rather than commodities. Although the contribution Hotelling made to utility theory is a matter of degree and sophistication in relation to the contribution of Marshall and Pigou, he succeeded in converting marginal cost pricing into a definite policy criterion, whereas the earlier writers recognized marginal cost pricing as an attractive possibility. Hotelling converted subsidy into a definite policy recommendation and set forth the type of taxes which should be used to raise the subsidy.

In a brief overview, marginal cost pricing is an outgrowth of welfare economics.[19] The policy of price equal to marginal cost is an outgrowth of the seven marginal conditions that characterize a welfare maximum defined either as (a) the maximization of consumers' surplus (Marshall) or (b) a situation in which no one can be made better off without making someone else worse off (Pareto).

The controversy that began with Hotelling's article pertains to the practical use of marginal cost pricing, a controversy that has continued

---

[18] Ibid., p. 242.

[19] A very excellent discussion on the evolution of marginal cost pricing is found in the following article: Ruggles, "The Welfare Basis of the Marginal Cost Pricing Principle," pp. 29-46.

for the past thirty years and remains unresolved. It appears that this long-standing controversy over marginal cost pricing prevented economic theory from providing an acceptable general rate theory. However, within the past decade, the issues, although still unresolved, are in a perspective which has contributed to progress at the applied level.

Limitations to Marginal Cost Pricing

Although marginal cost pricing has become a policy recommendation to be used when the private market mechanism fails to achieve the most efficient use of resources and the ideal output,[20] there are several instances where the use of marginal cost pricing is limited. Marginal cost pricing maximizes welfare under the assumptions that there are no joint costs, no relevant externalities, no distributional problems, independent demand functions, and infinite divisibility.[21] When these conditions are not satisfied, the solution derived from the application of marginal cost pricing must be modified.

Imperfect Competition

Marginal cost pricing is limited in its application when elements

---

[20]For detailed discussions see: William J. Baumol, Welfare Economics and the Theory of the State (2nd ed.; Cambridge: Harvard University Press, 1965), p. 107; Paul A. Samuelson, Foundations of Economic Analysis (New York: Atheneum, 1965), p. 247.

[21]Kafoglis, Welfare Economics and Subsidy Programs, p. 6.

of monopoly develop in the market. Monopolistic tendencies are found particularly on the selling side of the product market. In the case of monopoly, or imperfect competition, the firm maximizes profits by producing the output indicated by the equation of marginal revenue with marginal cost. The price exceeds marginal cost by an amount depending upon the elasticity of the demand function. Since price does not equal marginal cost when demand has any degree of inelasticity, the conditions necessary for the attainment of the optimum allocation of resources are not satisfied. If there are elements of monopoly in one or more sectors of the economy, the marginal conditions are not satisfied in any sector of the economy.[22] When one firm is required to follow a price equal to marginal cost policy, and the equality is not satisfied elsewhere, the firm tends to over produce with respect to the output of all other commodities. The basic problem resolves into one of achieving partial equality through marginal cost pricing and, therefore, attaining an output which more closely approximates the ideal output than that output produced when none of the marginal conditions is satisfied.[23]

Externalities

A condition which further complicates the attainment of efficiency through marginal cost pricing, or any form of market pricing for that

---

[22]Little, A Critique of Welfare Economics, p. 185.

[23]This argument is summarized in the following: Little, A Critique of Welfare Economics, pp. 162-65; Abba P. Lerner, The Economics of Control (New York: The Macmillan Co., 1944), pp. 134-36.

matter, is the existence of externalities--spillovers--which are not internalized in the free market. In these instances, positive government policy may encourage the attainment of the ideal output. The externality can take the form of either an external benefit or an external cost.[24] When significant external benefits are associated with the consumption or production of a commodity, the market, even if fully competitive, tends to produce too little of the commodity. When significant external costs exist, there is a tendency in the market to produce too much of the commodity. To achieve an optimal allocation of resources, it is desirable to expand the production of those commodities which have external benefits associated with them and contract the production of those commodities which have external costs connected with them. The means for implementing such policies are many and varied, but they are beyond the immediate scope of this study.

Decreasing Costs

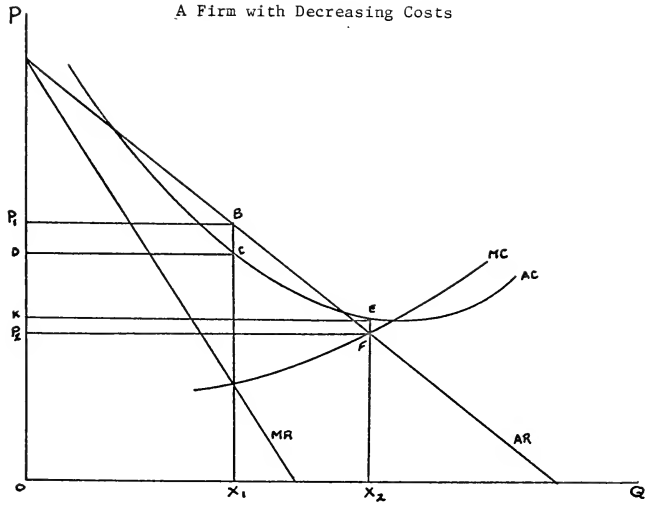
When a firm operates with decreasing costs, the use of marginal cost pricing converts a profit-maximizing firm into one incurring losses. The problem is demonstrated in Figure I below. If left to its own

---

[24]An external benefit is a benefit received from a user's consumption of a commodity by individuals other than the direct user. Everyone benefits from a vaccination received by any one member of the community since everyone's risk of contracting small pox is reduced. An external cost is a cost incurred from a user's consumption of a commodity and borne by individuals other than the direct user. An example is the social costs of caring for alcoholics resulting from the sale and use of liquor.

Figure I

A Firm with Decreasing Costs



devices, the firm produces output OX_1 where marginal revenue (MR) is equal to marginal cost (MC), and the selling price is OP_1. Profits are equal to the area P_1BCD. When marginal cost pricing is imposed, the firm is forced to equate marginal cost with price. To meet this requirement, the firm now produces output OX_2 and sells the output at the price OP_2. The firm now has losses equal to P_2KEF. The losses have been generated because the selling price (OP_2) is less than average cost (X_2E). With these losses, the firm will discontinue production in the long-run unless it receives some form of subsidy.

It is important to note two factors which are relevant in determining whether or not a subsidy is to be paid. The first factor is consumer demand as reflected in a collective context. Consumers must reveal a preference for the commodity to be produced and a willingness to pay the subsidy. The preference and willingness to pay criterion requires showing that the total benefits exceed the total costs to the community. When benefits exceed costs, consumer preferences should reveal a willingness of the consumers to pay the subsidy. The second factor is the cause of the losses. In developing a system of efficiency taxes and prices, the nature of the losses becomes a relevant factor. What might appear to be a firm operating with decreasing costs might actually be a non-optimum sized firm with increasing costs. In this instance, the losses result from improper decision-making by the owners of the firm. If losses arise because improper business decisions have led to excessive investment which created decreasing costs in the relevant range of output, a correct price structure would lead to overall losses. However, the losses are a correct penalty and should continue until the investment is brought down to proper size.[25] There is no need for a subsidy. For many reasons, this particular dictate has not been followed by public policy, especially in the case of railroads where externality, distribution, and politics have played an important role.

---

[25]William G. Shepherd, "Marginal Cost Pricing in American Utilities," Southern Economic Journal, XXIII (July, 1966), p. 60.

Interpersonal Comparisons and Equity

Marginal cost pricing has been attacked on the grounds that it fails to come to grips with the distributional problem.[26] Critics maintain that any decision about pricing has distributional implications and, therefore, imposes an interpersonal comparison of utility. Thus, it is maintained that marginal cost pricing assumes (explicitly or implicitly) that the income distribution generated by its application is superior, or at least not inferior, to the distribution that existed before it was applied. It can be demonstrated that the gains which accrue as a result of marginal cost pricing always exceed the losses, i.e., the gainers can more than compensate the losers. However, marginal cost pricing coupled with compensation seems impractical and would involve, for example, compensation to a monopolist upon the institution of policies designed to reduce his price to marginal cost.

The critics feel that the use of marginal cost pricing assumes that individuals who gain are more important, have more relevant tastes, or should be given more weight than the individuals who lose from the change. If it is in some sense "better" for the gainers to gain than it is for the losers to lose, a decision to adopt marginal cost pricing would seem to be in order. However, such a decision requires an interpersonal comparison. Such comparisons may seem reasonable if it is

---

[26] See: J. deV. Graaff, Theoretical Welfare Economics (Cambridge: Cambridge University Press, 1963), pp. 154-55.

established that the gainers have a greater propensity to enjoy than do the losers. The basis for making this judgment can be either economic or political.

> Any standard for the attainment of economic efficiency necessarily assumes the existence of some process for the measurement and aggregation of individual costs and utilities. If we define social welfare as a function of individual welfare, this process cannot be authoritarian but must reflect the voluntary choice of individuals. These choices can be expressed either through the market or through the polls. One gives weight to purchasing power, the other to political power.[27]

Advocates of marginal cost pricing usually avoid the problem of distributional decisions and interpersonal comparisons by assuming that the net gains are distributed in a manner that either adequately compensates the losers or does not lead to an inferior distribution of income. There is also the unlikely possibility that the gains are distributed randomly, leaving the distribution of income unchanged but at a higher general level. This latter possibility hinges on the assumption that there is no consistent institutional bias which favors some groups more than others in the distribution of gains--an assumption which seems unwarranted.[28]

---

[27]Kafoglis, Welfare Economics and Subsidy Programs, p. 7.

[28]For the development of this point see: Irving K. Fox and Orric C. Herfindahl, "Attainment of Efficiency in Satisfying Demands for Water Resources," American Economic Review, LIV (May, 1964), pp. 198-206.

In the case of distributional equity, the criticism misses the central point of the marginal cost pricing principle. It is true that marginal cost pricing generates an income distribution different from that which existed before its application. If the distribution is to be maintained, however, and the problem of redistribution is to be avoided, the principle of compensation can be used.[29] To maintain the original distribution of income, although possibly impractical, the gainers can always compensate the losers for the change (assuming away such problems as the monopolist who is compensated for reducing his price). When the sum of the benefits exceeds the losses, and compensation is paid to the losers, the aggregate effect is a net increase in welfare with no one worse off, and the change is consistent with the Paretian welfare criterion.

There are other adjustments which can be made to retain distributional equity. If the income distribution created through marginal cost pricing is not considered "desirable," an adjustment can be made

---

[29]For discussions of the compensation principle see: Abram Bergson, "A Reformulation of Certain Aspects of Welfare Economics," Quarterly Journal of Economics, LII (February, 1938), pp. 310-34; John R. Hicks, "The Foundations of Welfare Economics," Economic Journal, XLIX (December, 1939), pp. 696-712; Nicholas Kaldor, "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," Economic Journal, XLIX (September, 1939), pp. 549-52; Tibor Scitovsky, "A Note on Welfare Propositions in Economics," Review of Economic Studies, IX (1941-42), pp. 77-88.

through a lump-sum tax.[30] The lump-sum tax can be imposed to acquire the more desirable distribution of income without destroying the equality of the marginal conditions created through the use of marginal cost pricing. The proceeds from the lump-sum tax can be used as the source of revenue for compensation. The fixed plant costs can be paid by the state from revenues derived from the tax. The use of income or excise taxes, which impinge on resource margins, will destroy equalities in the marginal relationships. Income taxes will alter the equilibrium between work and leisure, and excise taxes the equilibrium of the marginal rates of substitution between commodities.[31] In her discussion of marginal cost pricing, Nancy Ruggles points to the key factor in determining the means of financing subsidies:

> Introducing the compensation would in effect mean that the revenue for subsidizing any given product would have to be derived from the people who consumed the product, and not from anyone else. To do this without violating the marginal conditions, the levy would have to fall on the consumers' surplus derived by the purchasers from the consumption of that specific product. It could not bear upon

---

[30] The first to suggest a lump-sum tax as a policy was Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," pp. 242-69. A multi-part tariff was also suggested by: R.H. Coase, "The Marginal Cost Controversy," Economica, XIII (August, 1946), pp. 169-82.

[31] These points are well developed by Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," pp. 242-69; Richard A. Musgrave, The Theory of Public Finance (New York: McGraw-Hill Book Co., 1959), pp. 136-54; Nancy Ruggles, "Recent Developments in the Theory of Marginal Cost Pricing," Review of Economic Studies, XVII (1949-1950), pp. 107-26.

> the marginal unit purchased by any consumer, so any form of per-unit tax would be inadmissable. A tax that must fall upon a specific product, but not upon the marginal unit, would of necessity yield a form of price discrimination.[32]

By combining marginal cost pricing with lump-sum taxes paid by the gainers, it is possible to maintain economic efficiency, compensate losers, and avoid interpersonal comparisons of utility.[33] Notwithstanding these possibilities, the distribution of gains due to policy changes ultimately involves the political process and hinges more on the distribution of political power than on the distribution of market power.

Joint Supply and Indivisibilities

In the case where a commodity is supplied within a capacity constraint, and/or in jointness with another commodity, marginal cost pricing can be difficult to apply.

When indivisibility of some input exists, marginal cost pricing can lead to an allocation of resources other than the optimal allocation.

---

[32] Ruggles, "Recent Developments in the Theory of Marginal Cost Pricing," p. 121.

[33] In support of marginal cost pricing and this point see: J.C. Bonbright, "Major Controversies as to the Criteria of Reasonable Public Utility Rates," American Economic Review, XXX (May, 1940), pp. 379-89; Emory Troxel, "Incremental Cost Determination of Utility Rates," Journal of Land and Public Utility Economics, XVIII (1942), pp. 458-67. However, it should be noted that lump-sum taxes provide an alternative means of correcting the income distribution although it may not always be a practical alternative.

In such instances where there is input-indivisibility, the use of marginal cost pricing can lead to profits or losses rather than to a breakeven solution as a result of the firm's being under built or over built with respect to demand. The specific solution is indeterminate.[34] The actual financial position of the firm is determined by the demand at the time the additional capacity is added. A plant which is under built will yield profits under marginal cost pricing. However, after additional capacity is added, the firm may operate with losses due to excess capacity, but there is a net gain in surplus. This result is caused by the indivisibility of the capital input. The indivisibility can make the sum of the prices less than marginal costs, and the firm will require a subsidy. To avoid the use of subsidies, the firm may be permitted to practice price discrimination, but this type of pricing becomes a question of distributional equity and violates efficiency criteria.

If the commodity is supplied in jointness with another, such as beef and hides, or peak and off-peak water,[35] the use of marginal cost pricing is complicated by the inability to identify the marginal costs of the separate products. Marginal cost of the product is no longer a

---

[34]Oliver E. Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisiblity Constraint," American Economic Review, LVI (September, 1966), p. 824.

[35]In these cases, the concept of jointness is not used in the traditional sense as the beef and hides case. The use of joint supply in the case of peak load problems is employed in the sense that the vertical summation of demand functions is necessary to determine an efficient solution.

usable criterion for determining prices.

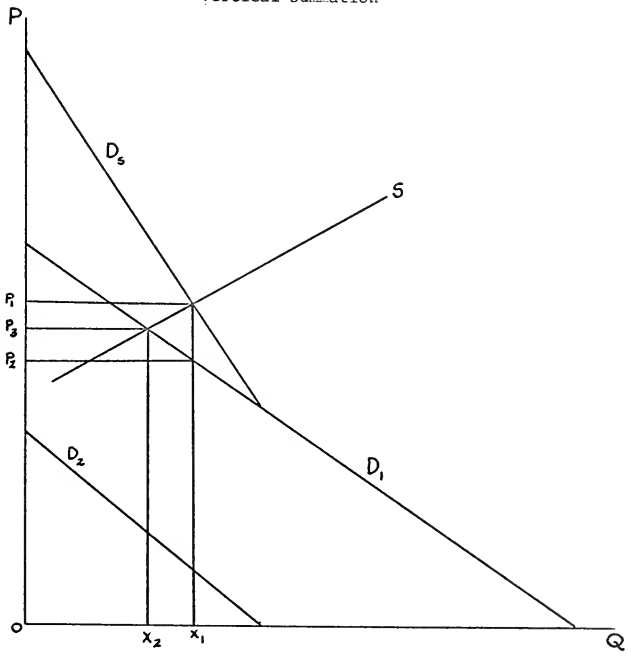
An Alternative Solution to Jointness, Externality, Decreasing Costs, and Indivisibility

The problems of externalities, decreasing costs, jointness, and indivisibilities have been mentioned as factors which impose restrictions upon the use of marginal cost pricing. The orientation of this study is toward municipal water utilities, and these problems have direct bearing upon an economic analysis of the rate practices of these utilities. A municipal water utility, in the process of providing water service, creates externalities in the form of improved public health and increased property values, and develops indivisibility in its capital plant, including jointness with respect to providing peak and off-peak water. Subsequently, these problems have an important bearing upon the solution which might be applied to municipal water rates.

A tool which has recently been developed in economic analysis is the vertical summation of demand functions. This tool has been applied separately to the problems of externalities and jointness. However, the power and general applicability of vertical summation as a geometric tool has not been fully recognized. This type of construction applies to the classic joint supply situation, to time jointness, and to external economies and diseconomies of consumption. Some of the possible interpretations are demonstrated in Figure II below where D_1 represents the demand for beef, D_2 the demand for hides, and S the supply function of cows. The optimal number of cows, beef, and hides is OX_1, and the Pareto optimal price is OP_2 for beef and P_2P_1 for hides. P_2 plus P_2P_1 equals OP_1 which, in turn, is equal to marginal cost. When

Figure II

Vertical Summation



two commodities are supplied jointly, the marginal cost of each commodity is not separable,and the price is based upon the demands for each of the two commodities. The price of each of the commodities does not equal the marginal cost of supplying it, but the sum of the prices is equated to the marginal cost of the joint output. Through the vertical summation of demands, the optimal output of both products can be

obtained where the sum of the prices is equal to marginal cost.[36] This solution apparently satisfies the criteria for economic efficiency.[37] It is noteworthy, however, that a redistribution of income which alters the demands will also alter the price relationships of the joint products, even under constant costs. This phenomenon will not occur in the case of separable products, and, thus, it places joint-cost pricing in a vaguely defined "in-between" area with respect to the distinction between equity and efficiency.

The analysis in Figure II can also be applied to time jointness if the diagram is reinterpreted so that D_1 is a peak demand, D_2 is an off-peak demand, and S is the cost of the facility which is available to serve both demands. The optimal quantity is OX_1 with a peak price of OP_2 and an off-peak price of P_2P_1. Again, at the output OX_1, the sum of the prices is equal to the marginal cost of providing the combined services to meet the peak and off-peak demands.[38]

---

[36]Howard R. Bowen, Toward Social Economy (New York: Rinehart and Co., Inc., 1948), pp. 177-80.

[37]Kafoglis, Welfare Economics and Subsidy Programs, pp. 21-33.

[38]The development of time jointness has been notable in the theory of public utility rate structures because of the inability of the utility to adjust its capacity to meet the peak and off-peak demands. If the utility could adjust its plant size neatly to these demands, a joint cost problem would not exist. However, it has been noted that the capacity required to produce the output for one demand automatically provides the capacity to produce for the other demand. The capacity to produce water at four o'clock is the same capacity used to produce water at eight o'clock. Although the outputs in the two periods are different, the capacity is the same, and this fixed proportion of capacity between the two periods is the basic requirement for a joint cost problem. See:

If D_2 is interpreted as the value of the "spillover" at the margin to individual A as a result of individual B's consumption, A is an indirect beneficiary. Optimal consumption for B is not the quantity OX_2 at price OP_3, the amount determined by the market; it is OX_1, the amount forthcoming only if some means is found to reduce the price to individual B below the supply price (marginal cost). This interpretation applies to vaccines and other quasi-collective goods, where it is necessary for individual A to guarantee his own health by subsidizing individual B's consumption through public subsidy or other means requiring governmental action.[39]

The same type of analysis is sometimes applied in the cases of excess capacity and indivisibility where marginal cost pricing leads to losses. The overhead (the loss) is considered a collective, or joint, input which applies to both demands. However, a fixed overhead allocated

---

Donald H. Wallace, "Joint Supply and Overhead Costs and Railway Rate Policy," Quarterly Journal of Economics, XLVIII (August, 1934), pp. 583-616. The joint supply problem is a difficult one to resolve as evidenced by the "Pigou-Taussig controversy," and many of the issues still remain unclear. The inability to adjust capacity to the peak and off-peak demands in the same time period has been treated as a joint cost problem by several notable writers. See: James M. Buchanan, "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXX (August, 1966), pp. 463-471; M.A. Crew, "Peak-Load Pricing and Optimal Capacity: Comment," American Economic Review, LVIII (March, 1968), pp. 168-70; Peter O. Steiner, "Peak Loads and Efficient Pricing," Quarterly Journal of Economics, LXXI (November, 1957), pp. 585-610; Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," pp. 810-27; William S. Vickery, Microstatics (New York: Harcourt, Brace and World, Inc., 1964), pp. 225-44.

[39]This analysis assumes that all individuals will reveal their true preferences.

through vertical summation, in these instances, does not lead to a Pareto efficient solution. In the case of common costs, the demands must be summed horizontally, and each user pays the same price and contributes to the overhead in proportion to the total output that he takes.[40]

The type of analysis embodied in Figure II may be applied to the entire range of congestion problems, multi-product problems, and quasi-collective goods. It is possible to analyze many problems through a simultaneous application of collective demands (vertical summation) and the marginal cost standard. These constructions are crucial to the application presented in Chapters V and VI.

Interdependent Demands

Marginal cost pricing assumes that individual demands are independent of each other. When demands are not independent, marginal cost pricing weakens. For example, in the case of water, the hourly peak demand is a partial function of the hourly off-peak price and the off-peak demand is a partial function of the peak price. The interdependence of demands leads to several complex problems such as a shifting peak and the need for continual price adjustments when there is a price differential between the peak and off-peak periods (This point is developed further in the following chapter.). These problems are both administrative and theoretical. Consequently, they involve a highly

---

[40] See: Wallace, "Joint Supply and Overhead Costs and Railway Rate Policy," pp. 583-616.

technical type of analysis to determine their solution. However, a combination of fixed charges and marginal cost pricing provides a means of overcoming these problems. The solution is developed in Chapter V.

Conclusions

Many of the objections raised about marginal cost pricing have rested upon theoretical grounds. In general, the criticisms have been aimed at the underlying assumptions of the marginal cost pricing principle. To the extent that the assumptions are not fulfilled, the marginal cost pricing principle is limited as a single pricing guide. However, these limitations do not mean that marginal cost pricing has no validity as a useful pricing guideline. It has been demonstrated that marginal cost pricing, combined with vertical summation of demand curves, permits the determination of a Pareto efficient solution in cases of joint products, joint costs, externalities, and indivisibilities. Marginal cost pricing provides a take-off point as a basis for determining the rate structure and the size of plant necessary to achieve the ideal output.[41]

Some of the most damaging criticism of marginal cost pricing has been practical. In many respects, the practical difficulties have been the

---

[41]William S. Vickery, "Some Implications of Marginal Cost Pricing for Public Utilities," American Economic Review, XLV (May, 1955), pp. 605-20.

deterrents which have kept economic theory from making a generally acceptable contribution to utility rate theory until the last decade.[42] Some critics maintain that industrialists do not think in terms of marginal costs. This point has not been established. The industrialist may well have his own terminology for what the economist labels marginal cost. The inconsistency in terminology between the economist and the industrialist is not a valid basis for discarding marginal cost pricing.[43]

The other practical objection to marginal cost pricing relates to cases of joint supply. Because of the difficulty in identifying marginal costs, some critics advocate the use of average cost pricing as an alternative.[44] This argument throws the baby out with the water. If it is difficult to identify marginal costs, it is equally difficult to identify average cost. In either case, the costs must be allocated between the

---

[42]Kenneth E. Boulding, Economic Analysis: Microeconomics (New York: Harper and Row, Publishers, 1966), pp. 498-99.

[43]For a discussion of this problem see: Fritz Machlup, "Theories of the Firm: Marginalist, Behavioral, Managerial," American Economic Review, LVII (March, 1967), pp. 1-33.

[44]Boulding, Economic Analysis: Microeconomics, pp. 498-99; Wallace, "Joint Supply and Overhead Costs and Railway Rate Policy," pp. 583-616.

commodities,and average cost pricing does not eliminate the problem.[45]

[45]For excellent discussions of all the major aspects of marginal cost pricing and the controversy over its usefulness see: J. deV. Graaff, Theoretical Welfare Economics, pp. 142-55; Little, A Critique of Welfare Economics, pp. 185-216; Ruggles, "Recent Developments in the Theory of Marginal Cost Pricing," pp. 107-26; William S. Vickery, "Some Objections to Marginal Cost Pricing," Journal of Political Economy, LVI (June, 1948), pp. 218-238.

CHAPTER III

THE THEORETICAL MODELS

Introduction

During the past decade, many writers have advocated a return to the marginal cost pricing principle as a basic policy recommendation in the field of public utilities, and significant contributions to utility rate theory have been made. One of the more significant contributions of recent writers is their handling of the difficult peak load problem which entails the problem of time jointness with respect to supply. Taken in their entirety, these contributions form a relatively complete analysis of the problems faced in determining optimal capacity and optimal rate structures.[1] The purpose of this chapter is to evaluate these models in terms of their analyses, assumptions, and distributional implications in order to determine their applicability to the special problems of municipally owned water utilities.

---

[1]The models included in this chapter are: James M. Buchanan, "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXX (August, 1966), pp. 463-71; Ralph K. Davidson, Price Discrimination in Selling Gas and Electricity (Baltimore: Johns Hopkins University Press, 1955); Jack Hirshleifer, "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXII (August, 1958), pp. 451-62; Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica, VI (July, 1938), pp. 242-69; Peter O. Steiner, "Peak Loads and Efficient Pricing," Quarterly Journal of Economics, LXXI (November, 1957), pp. 585-610; Oliver E. Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," American Economic Review, LVI (September, 1966), pp. 810-27.

The Hotelling Model

One of the first rigorous solutions to advocate marginal cost pricing was that presented by Harold Hotelling.[2] Hotelling attempted to develop pricing policies which would lead to maximum consumers' surplus. Through the use of mathematics, Hotelling rigorously demonstrated that the optimum for the general welfare corresponds to the sale of everything at marginal cost prices.[3] He refutes the position taken by utility engineers that commodities produced by industry must be sold at a price high enough to cover full costs. Indeed, he states that this policy leads to economic inefficiency and to a loss of welfare as measured by consumers' surplus.[4] The essence of Hotelling's analysis is summarized in Figure III below. Assuming constant short-run marginal costs, a capacity constraint at output OX_o, and demand as shown by D_1, the welfare maximizing firm supplies the output OX_1, at the price OP_1, which is equal to short-run marginal cost (b). As demand increases, the firm expands output toward OX_o at a constant price (OP_1) until output OX_o is reached. As demand continues to increase, the firm cannot expand output beyond OX_o, except in the long-run. In the short-run, the price moves upward along the vertical segment NR. At price OP_2, with demand now at D_2, the firm is charging a price which

---

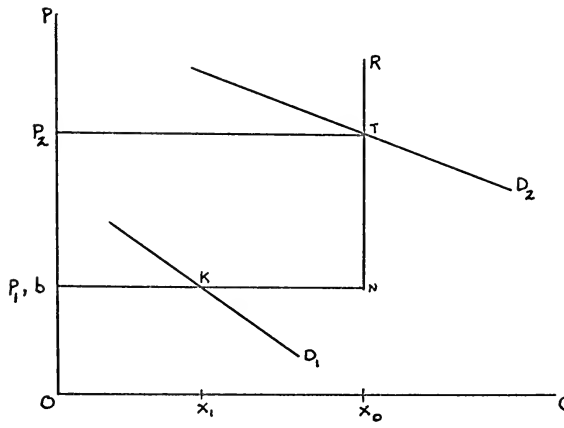
[2]Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," pp. 242-69.

[3]Ibid., p. 242.

[4]Ibid., p. 242.

Figure III

The Hotelling Model



exceeds short-run marginal production costs by an amount equal to NT. The surplus NT of price over marginal cost is the effect of rationing a fixed supply through the use of efficiency pricing. Hotelling's position on this surplus is twofold. At one point he defines the surplus as a rental charge not unlike the site rental of land.[5] Such a rental charge, according to Hotelling, becomes a source of revenue to the state and can be taxed away without affecting the allocation of resources. However, the charge is also needed to reflect the "social costs" of congestion. As the quantity demanded exceeds the capacity

[5] Ibid., p. 249.

to produce, the utility becomes overcrowded, thus reducing the quality of service to all users, and Hotelling feels that the price should be high enough to reflect these social costs of congestion. The interpretation Hotelling gives to the vertical segment NR is a marginal opportunity and/or a marginal social cost; not a marginal money cost of production as the segment bN reflects. Prices OP_1 and OP_2 are welfare maximizing.

The basic elements of this model provided the background and the framework upon which more recent writers have based their analyses. The change in the price policy from a situation such as that defined by point K in Figure III to that defined by point T provided the basis for the eventual development of sophisticated solutions to the peak load pricing problem.

The Steiner Model

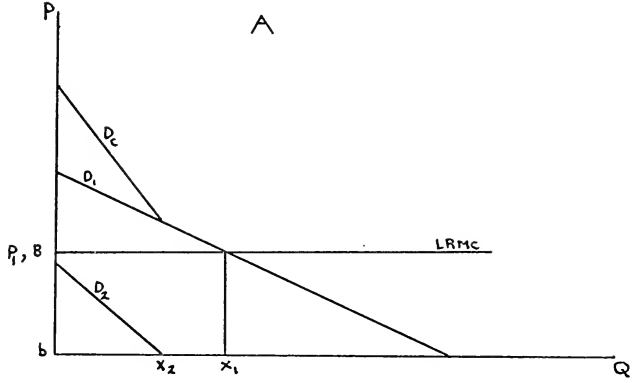
One of the first general solutions to the problem of defining an optimal plant and rate structure for firms facing a peak load problem was presented by Peter O. Steiner.[6] Subsequent models have been variations of and improvements upon Steiner's analysis. The proper policy,

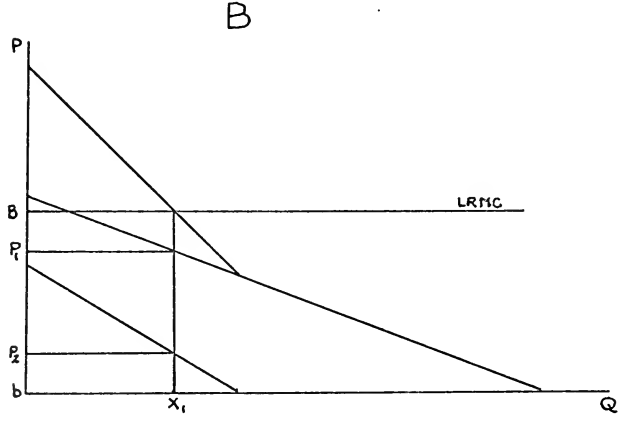
---

[6]Steiner, "Peak Loads and Efficient Pricing," pp. 585-610. At the time the Steiner article appeared, a similar solution appeared in France. See: Marceo Boiteux, "La Tarification des Demandes en Pointe: Application de la Théorie de la Vent au Coût Marginal," trans. by H.W. Izzard, Marginal Cost Pricing in Practice, ed. by James R. Nelson (Englewood Cliffs: Prentice-Hall, 1964), pp. 59-89. It should be noted that the nature of the solution was recognized by transportation economists many years ago. See: Michael R. Bonavia, The Economics of Transport (New York: Pitman Publishing Corp., 1936), pp. 103-11.

FIGURE IV

The Steiner Model





according to Steiner, is one which achieves a social optimum, i.e., the policy which maximizes the excess of expressed consumer satisfaction over the cost of the resources used in production.[7] Steiner's model attempts to formulate a price policy which leads to the optimal amount of physical capacity and which is consistent with marginal social costs.

Using a vertical summation of peak and off-peak demand curves, Steiner presents the framework shown in Figures IV_A and IV_B where (b) --the horizontal axis-- is the short-run marginal operating cost (assumed to be zero), and bB represents the long-run marginal capacity costs. In Figure IV_A, there is excess capacity (X_2X_1) during the off-peak period (D_2), and the off-peak user places no demand on the system capacity. The off-peak user pays a price equal to the short-run marginal operating costs (b). In Figure IV_B, the solution entails no excess capacity during the off-peak period. Both the peak user and the off-peak user make a contribution to the capacity costs. The off-peak user pays the price bP_2, the peak user pays the price bP_1, and the sum of the two prices is equal to the marginal capacity costs (bB). Steiner's general argument is that a unit of capacity can be added if the costs can be covered by the sole demand of any one period (D_1 in Figure IV_A) or by the combined demands of two or more periods (D_1 plus D_2 in Figure IV_B).[8]

---

[7] Steiner, "Peak Loads and Efficient Pricing," pp. 585-87.

[8] Ibid., p. 589.

Upon arriving at the solution in Figure IV_B, Steiner concludes that an efficient pricing solution entails discrimination. In Figure IV_B, for example, the off-peak and peak outputs are equal, but the prices are unequal. Steiner states:

> If demand curves are different, at a given output, the prices are unequal and since this is truly a case of joint costs, unequal prices in the face of equal output and joint costs means discriminatory prices.[9]

This statement reflects Steiner's recognition that in the case of a firm having both a peak load problem and a capacity constraint (short-run) the peak load problem becomes one of joint costs. Prices do not equal short-run marginal operating costs, but they maintain long-run optimal capacity. The strength of the relative demands determines each period's users' share of the utility's capacity costs. Although the prices are discriminatory in the usual sense, the pricing solution is efficient in the Paretian sense and satisfies the criteria of welfare maximization. The sum of the prices (P_1 plus P_2) is equal to the long-run marginal costs (Bb). Therefore, Steiner's formulation does satisfy the criteria for welfare maximization.

The Hirshleifer Model

An alternative formulation of the peak load problem has been developed by Jack Hirshleifer, who employs a different concept of marginal

---

[9] Ibid., p. 590.

cost than Steiner, which leads him to reject Steiner's description of the optimal solution as one which involves price discrimination.[10] The difference between the peak and off-peak prices is explained in terms of marginal opportunity costs.

Hirshleifer uses the same optimizing criteria as Steiner, but marginal cost is defined as a marginal opportunity cost when the utility faces a capacity constraint. The marginal opportunity cost is the value set upon the resources in the most valuable alternative use being sacrificed.[11] Hirshleifer divides constant long-run marginal costs between joint and separable long-run cost elements. The joint long-run marginal cost is the cost per-combined-unit of production--a variable cost for each of the two periods plus a capacity cost. The separable long-run marginal cost is the cost of increasing the output of one of the two periods, the output for the other period being held constant.[12]

Hirshleifer's solution is based upon two different short-run marginal cost functions. One solution assumes a continuous cost function, the other a discontinuous cost function. Hirshleifer's model, based upon the assumption of constant long-run costs and a continuous short-run marginal cost function, is shown in Figure V below. Each period's price is equal to short-run marginal cost. The off-peak price is OC and the

---

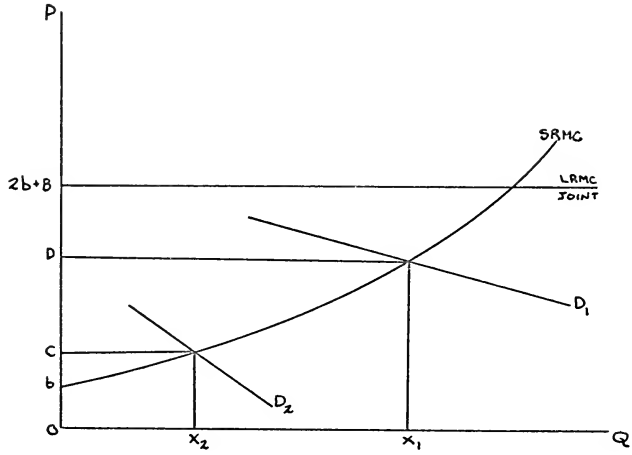
[10] Hirshleifer, "Peak Loads and Efficient Pricing: Comment," pp. 451-62.

[11] Ibid., p. 451.

[12] Ibid., p. 455.

Figure V

Hirshleifer's Continuous Cost Model



quantity supplied is OX_2. The peak price is OD and the quantity supplied is OX_1. The short-run conditions for welfare maximization are satisfied. Since the short-run marginal cost function is continuous and each period's demand is equated with this cost function, there is no discrimination in the pricing solution. Each period's price is equal to short-run marginal production costs.

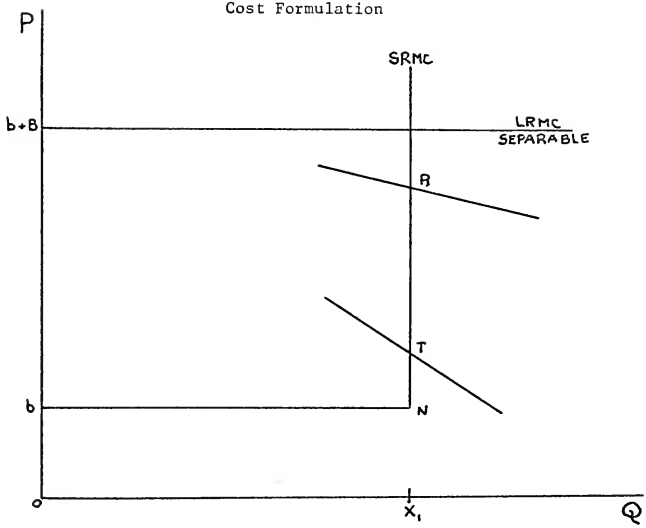
The long-run conditions for welfare maximization are also satisfied. The sum of the prices (OC plus OD) is equal to long-run marginal costs (2b plus B). If the sum of the prices is not equal to long-run marginal cost, total surplus can be increased by restoring the equality. If the sum of the prices is greater than long-run marginal costs, an

addition to capacity increases consumers' surplus by more than producers' surplus is reduced. If the sum of the prices is less than long-run marginal costs, producers' surplus can be increased by a greater amount than consumers' surplus is reduced by a contraction of capacity.

Steiner's conclusion about discriminatory prices was based upon the assumption of joint supply combined with a discontinuous short-run marginal cost function. Hirshleifer also recognizes the problem of a vertical short-run marginal cost function, but he does not use the joint cost assumption. According to Hirshleifer, the vertical segment leads to an indeterminate marginal cost as a cash outlay concept. Once the maximum output is reached, short-run marginal cost becomes indeterminate. Hirshleifer's formulation on the assumption of a discontinuous marginal cost function is shown in Figure VI. In the framework below, X_1T is the off-peak price, and X_1R is the peak price. The sum of the prices (X_1T plus X_1R) is equal to long-run marginal cost (b plus B). The quantity OX_1 is supplied during both the peak and off-peak periods. Once output OX_1 is reached, marginal cost increases from Ob (operating costs) to Ob plus B (long-run marginal cost). Short-run marginal cost is indeterminate along the vertical segment beginning at point N. Although two different prices are charged for the same output, Hirshleifer explains the difference in terms of costs. He departs from the marginal money cost concept and explains the difference in terms of the differences in the marginal opportunity costs between the two periods; therefore, his solution described above does not entail price

Figure VI

Hirshleifer's Discontinuous
Cost Formulation



discrimination.[13] This approach puts Hirshleifer in a position

[13]Crucial to his analysis is Hirshleifer's explanation as to why this argument does not explain away all forms of discrimination. Hirshleifer explains that in the textbook cases of discrimination "...the market is divided artificially. The commodity being the same, at the profit-maximizing solution, the marginal customers in each class, while paying different prices, are being served at the same opportunity cost--the value of the first unit of unsatisfied demand in the higher priced market is the most valuable alternative foregone. Therefore, no price difference is justified on opportunity cost grounds. In the case under consideration, the market division is not artificial--taking a unit away from the off-peak does not make it possible to supply a unit on-peak, so the higher on-peak value is not the relevant alternative social opportunity cost of the off-peak service," ibid., p. 459.

different from that of both Steiner and Hotelling. As mentioned earlier, Hotelling interprets the price differentials in terms of the marginal social congestion costs and the ensuing deterioration of service resulting from congestion. Steiner treats the problem as a joint cost problem. Hirshleifer, on the other hand, defines the price differentials in terms of marginal opportunity costs.

Hirshleifer concludes that, when the marginal cost function is continuous, prices are not discriminatory since they are equal to short-run marginal money costs. When the marginal cost function is discontinuous, Hirshleifer concludes that prices are not discriminatory in the Steiner sense since prices are equal to marginal opportunity costs. In this context, Hirshleifer's definition of marginal cost is not the traditional money outlay concept of marginal cost, but the opportunity cost interpretation, and the marginal value of the last unit is equal to the value of the first unit of unsatisfied demand in each of the periods. Therefore, prices equated to opportunity costs are determined by the relative strength of the demand functions in each of the periods. With the exception of interpretation, both solutions are, for all practical purposes, identical.

The Williamson Model

A recent development in the peak load pricing problem has been the framework developed by Oliver E. Williamson.[14] The Williamson model was

---

[14]Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," pp. 810-27.

presented to improve upon the geometry developed by Steiner and the concepts employed by Hirshleifer. Williamson recognizes that, for electric utilities (equally valid for water utilities), the peak period is of longer duration than the off-peak period. When the two periods are of unequal duration, an adjustment must be made in the vertical summation since only an assumption of equal time periods permits straightforward summation such as the Steiner summation. This adjustment is the central point of Williamson's geometric model. When the two periods are of unequal duration, a weighted vertical summation of the two demands is required.

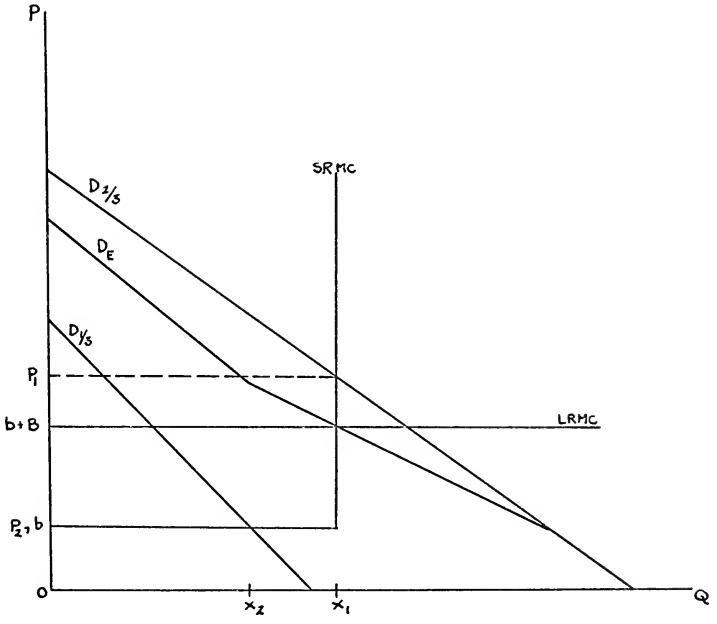
Williamson's model, shown in Figure VII below, assumes constant costs, a peak period of sixteen hours, and an off-peak period of eight hours. The measure of welfare is total surplus as developed by Marshall under an assumption of constant costs. Total welfare is equal to (total revenue + consumers' surplus) minus total costs.[15] In Figure VII, the short-run solution equates each period's price with the short-run marginal cost. The peak price is OP_1 and the off-peak price is OP_2. The long-run solution equates long-run marginal cost with the effective demand curve D_E, which is the weighted sum of the individual demands. Since it is assumed that the peak lasts for sixteen hours and the off-

---

[15]For the technical application of consumers' and producers' surplus see: John R. Hicks, Value and Capital (2nd ed.; Oxford: Oxford University Press, 1946), pp. 38-41; John R. Hicks, Revision of Demand Theory (London: Oxford University Press, 1956), pp. 67-106; John R. Hicks, "The Four Consumer Surpluses," Review of Economic Studies, XI (1943), pp. 68-74; John R. Hicks, "The Generalized Theory of Consumer's Surplus," Review of Economic Studies, XIII (1945-46), pp. 68-74.

Figure VII

The Williamson Model



peak for eight hours, the peak demand is given a weight of 2/3 and the off-peak a weight of 1/3. The weighted curve reflects the average price the utility receives over the entire cycle which can be applied toward the capacity costs. It is basically a "long-run" demand curve.

Using Williamson's model, the basic conclusions are threefold. First, optimal price in every subperiod is given by the intersection of the short-run marginal cost and the subperiod demand. Secondly, plant size

is given by the intersection of the effective demand for capacity curve and the long-run marginal cost function. Thirdly, in a fully adjusted, continuously utilized system with only two period loads, (a) peak load price always exceeds long-run marginal cost, (b) off-peak price is always below long-run marginal cost, and (c) only when the off-peak fails to utilize capacity when priced at short-run marginal cost does the peak load bear the entire burden of the capacity costs.[16]

As in the case of the Steiner and Hirshleifer models, the distribution of the capacity costs depends upon the relative strengths of the two demands when prices are equated with short-run marginal cost. When plant is divisible, making optimal capacity possible, the utility operates with zero net revenues in the long-run. The surplus generated during the peak period exactly equals the deficit experienced during the off-peak period.

Williamson modifies the assumption of a completely divisible plant in his model (an assumption implied in the other models) by constructing a solution based upon the alternative assumption of indivisible plant. In this model, there is no guarantee that revenues will be sufficient to cover costs. Both the peak and off-peak prices can be below long-run marginal cost because of the indivisibility of the capital stock. Since prices are not sufficient to cover the long-run marginal costs, the firm

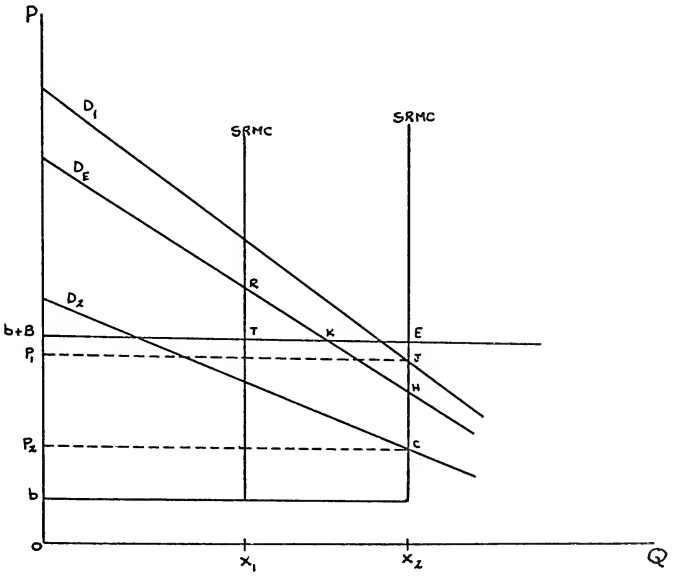
---

[16]Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," pp. 821-22.

operates with negative net revenues. The solution is shown in Figure VIII below, where the peak price is OP_1, and the off-peak price is OP_2. It can be seen that the off-peak price and the peak price are both less than long-run marginal cost, and the firm is operating with total losses equal to (P_2CEb+B) during the off-peak period plus (P_1JEb+B) during the peak period.

Figure VIII

The Williamson Model
Assuming Indivisibility



Although the firm is operating with losses, the plant size is optimal from a welfare standpoint because the effective demand and the long-run marginal cost function are the co-determinants of plant size. When indivisibility is present, the firm operates with zero net revenues only accidentally. In Figure VIII, assume output OX_1 is the present capacity, and D_1 and D_2 are the peak and off-peak demands respectively. Based upon these demands, output and capacity should be that defined by point K where the effective demand intersects long-run marginal costs. Because of the indivisibility, the firm must add capacity which is capable of producing output OX_2. The addition to capacity is warranted, however, because the gain in consumers' surplus exceeds the loss in producers' surplus, as shown by the relationship between the two triangles RKT and EKH. The triangle RKT is larger than EKH so there is an increase in total welfare from the expansion of capacity from OX_1 to OX_2, with the net increase equal to the difference between these two triangles. It is evident that there is no guarantee that the number of units of output capable of being produced is the exact number needed to put the firm in a position to realize zero net revenues when indivisibility exists. The result might be profits, losses, or breakeven.

The general conclusion of Williamson's model is that in the case of divisibility, a plant size is optimal if "...an increase in scale leads to a decrease in producers' surplus that exceeds the gain in consumers' surplus; a decrease in scale yields an increase in producers' surplus that is less than the loss of consumers' surplus."[17]

---

[17] Ibid., p. 820.

The Davidson Model

Another work representing a variation on the former models, but which came earlier in time, is the model formalized by Ralph K. Davidson.[18] Davidson's study was a significant and comprehensive contribution to the practical application of discriminatory prices. In the course of a more general analysis, Davidson derives a solution to the peak load problem. The significant feature of this solution is the use of the long-run as the relevant time period for policy determination. Davidson argues that prices equated with long-run marginal costs are more relevant for policy purposes, whereas the other models equate price with short-run marginal cost. Davidson also concludes that a rate schedule should not be discriminatory, and that all rate differentials should be based upon costs. Although Davidson's work is frequently cited in other works on utility rate theory, it can be demonstrated that, in reaching his conclusions, Davidson ultimately develops an average cost pricing solution.

Davidson's basic analysis is summarized in Figure IX below. Each period's price is equated to the long-run marginal cost of supplying the period. The difference between the long-run marginal cost functions reflects the cost differences of serving the two periods. As plant is expanded to meet the peak demand (D_1), the firm has greater operating and capacity costs. To expand production during the off-peak (D_2), the

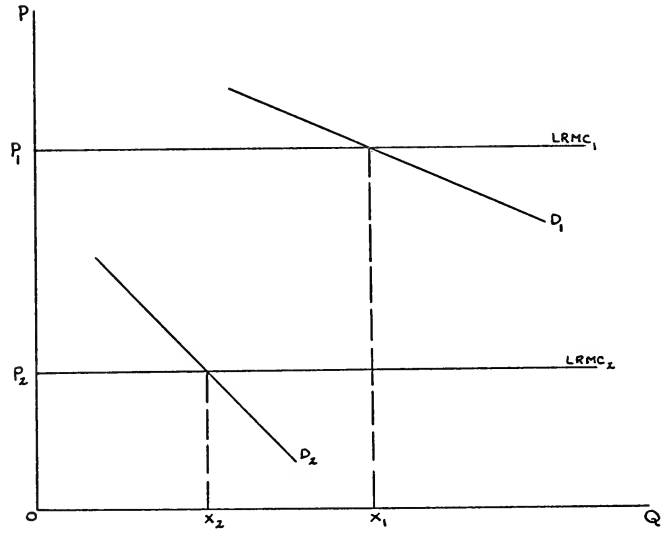
---

[18] Davidson, Price Discrimination in Selling Gas and Electricity (Baltimore: Johns Hopkins University Press, 1955).

utility does not have to expand capacity. The only additional costs are operating costs. According to Davidson, it is this solution which maximizes total surplus.

Figure IX

The Davidson Model

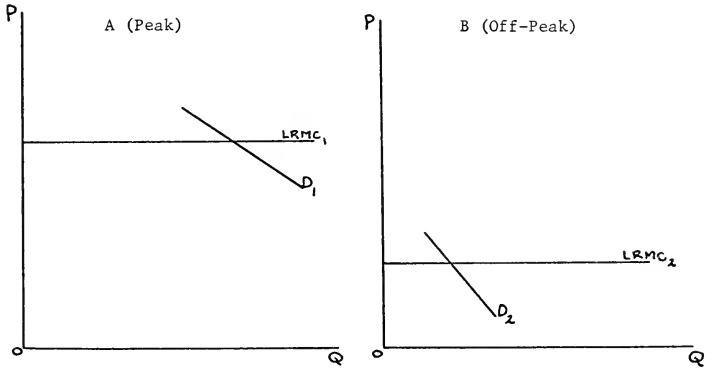


It is difficult to understand why a firm's long-run marginal cost function can be high or low depending upon which period's production the firm attempts to expand. Davidson has implicitly separated the time periods into separate markets. The off-peak period is one market

and the peak period is the other. Each of the markets has its own long-run marginal cost function. The analysis is presented in Figure X below.

Figure X

An Interpretation of
the Davidson Solution

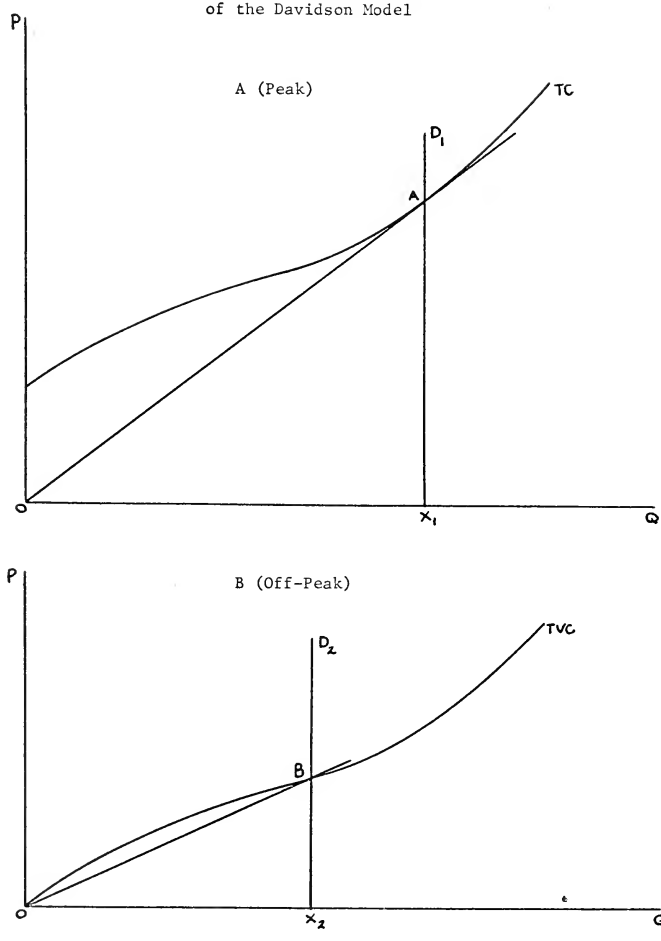


Plant size is sufficient to meet the peak demand (the determinant of plant capacity), and the off-peak production is, therefore, always less than capacity.

Davidson fails to consider that a single firm is producing both the peak and the off-peak outputs with the same basic plant capacity. Figure XI below demonstrates the implicit analysis of the Davidson model. The capacity of the plant is determined by the expected peak demand (D_1). To achieve economic efficiency, the plant operates at the minimum point of the short-run average cost function (given Davidson's assumption of

Figure XI

A Total Cost Interpretation
of the Davidson Model

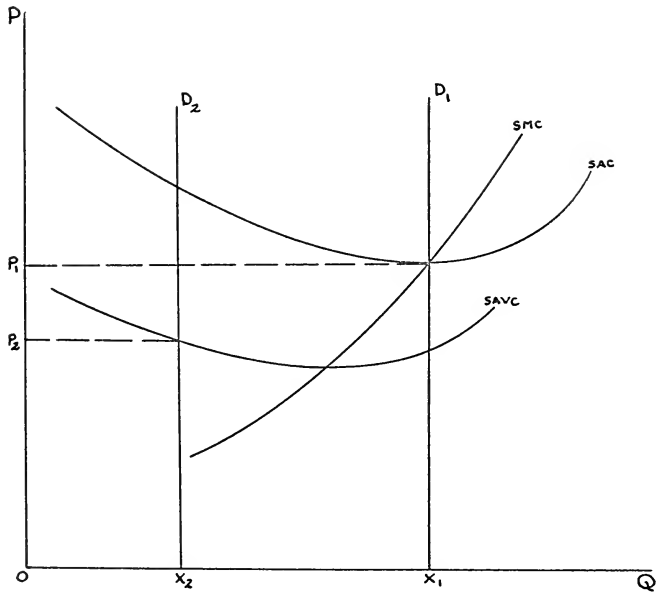


constant long-run average and marginal costs). During the peak, the utility produces output OX_1 with total costs of X_1A. The peak price is X_1A divided by OX_1. Peak production is carried to the point where short-run average cost equals short-run marginal cost equals long-run marginal cost. It appears that Davidson selects long-run marginal cost rather than short-run average cost to equate with price since they are equal to each other. Therefore, price is also equal to short-run average cost. Off-peak price is determined by dividing total variable costs of X_2B by the off-peak output of OX_2. Since capacity costs have been recovered during the peak period, variable (operating) costs are the only relevant costs for the off-peak period.

The Davidson solution can also be presented by using average cost curves shown in Figure XII. Expected peak demand determines the size of plant. If the firm is operating under constant cost conditions, economic efficiency is achieved when short-run average cost is a minimum at the peak output. The peak price is OP_1,and all the capacity costs (the difference between SAC and SAVC) are recovered from the peak period output. Variable costs are the relevant costs for the off-peak period. With an off-peak demand of D_2, the off-peak price is OP_2, and the only costs included in this price are the firm's variable costs. Davidson's pricing technique is entirely a cost-of-service based pricing solution, and it is the same, in most respects, as the solutions proposed by utility engineers and managers.

Figure XII

The Average Cost Interpretation
of the Davidson Model



That Davidson uses average cost pricing is further demonstrated by his definition of marginal cost.[19] He takes the position that the utilities must cover their long-run marginal costs if they are to have successful financial operations without having to resort to price

[19]It would appear that part of Davidson's misinterpretation also stems from his apparent failure to either recognize or acknowledge the jointness between the peak and off-peak periods.

discrimination.[20] Davidson cites the tendency of people in the utility field to use the term "total increment costs" to refer to long-run marginal costs.[21] Davidson uses this utility concept of incremental cost and calls it long-run marginal cost. The incremental cost interpretation permits the utility to have different long-run marginal cost functions depending upon the time period being served. Incremental costs are different depending upon whether output is expanded during the peak or the off-peak period. Davidson's concept of marginal cost is different from the economic concept of long-run marginal cost. The latter designates the change in costs associated with changes in production when all inputs are changed. Davidson's use of long-run marginal cost as incremental costs in which either all or only part of the inputs are changed depending upon the time period in question is evidenced by the following statement:

> The relevant cost concept consists of long-run marginal customer costs, and long-run marginal output costs, which always includes energy costs and may or may not include capacity costs depending upon the time of day and season of the year when the marginal unit is used.[22]

All inputs are changed during the peak period, and,during the off-peak period, the capital input is held constant. The use of incremental costs can lead to an average cost pricing solution.

---

[20] Davidson, Price Discrimination in Selling Gas and Electricity, p. 72.

[21] Ibid., p. 72.

[22] Ibid., p. 72.

The Buchanan Model

James M. Buchanan provides an analysis of the peak load problem, which represents a further sophistication and modification of the previous models.[23] Buchanan's contribution is the introduction of first degree discrimination into the standard models which destroys their apparent determinacy. According to Buchanan, the determinacy is "...produced only by the implicit adoption of unjustified assumptions concerning the uniformity of marginal price over quantity."[24] Buchanan maintains that there is no reason for a utility to charge the same price for all quantities demanded in each period by the same buyer. Buchanan's solution takes into consideration the block-rate structures that are actually used, and his position is summarized in the following statement:

> If the model should be restricted to goods, and not applied to services that are consumed as purchased, then the possibility of interpersonal resale within periods might tend to insure against the possibility of "price discrimination" over quantities sold to single buyers. However, it seems plausible to expect that "price discrimination" over quantities sold to single buyers will accompany "price discrimination" among separate buyers.[25]

The utility can charge different prices for different units sold to a single buyer. The price paid for the marginal unit is affected by

---

[23] Buchanan, "Peak Loads and Efficient Pricing: Comment," pp. 463-71.

[24] Ibid., p. 462.

[25] Ibid., p. 465.

the prices charged for the infra-marginal units. In this event, it becomes necessary to distinguish between marginal and average prices. For each marginal price, there may exist several quantities demanded as the price of the entire offer is altered. In these instances, no precise demand curve can be derived unless the buyer is presented with all the possible price offers. When a precise demand cannot be identified, vertical summation is impossible,and the solution becomes indeterminate.[26]

Buchanan reiterates the condition necessary to achieve Pareto optimality in joint cost situations: the sum of the prices at the margin must equal marginal cost. The use of discriminatory pricing in the infra-marginal units, however, affects the distribution of the cost shares among the various users. Buchanan maintains that the distribution of the cost shares over the infra-marginal units and the income effects of the discrimination can affect the location of the margin which, in turn, determines the marginal capacity and the marginal prices.[27] Through the manipulation of discriminatory rate structures, the utility can arrive at a set of marginal prices where the sum is equal to long-run marginal cost,and there is no excess capacity during the off-peak period. Since the rate structure only affects the cost shares, there is no significant effect on Pareto optimality. There are, however,

---

[26] Ibid., pp. 465-66.

[27] Ibid., p. 466.

significant distributive effects. Buchanan states that "...the choice among different price offer sets finally rests on the decision maker's evaluation of different distributions of consumers' surplus among the separate period's demanders."[28]

Buchanan has set up a unique model which reconciles price discrimination, distributional effects, and Pareto optimality. Once the price offer is given, the margin is located, and the plant size becomes determinate. The selection of the rate is made by the utility manager whose decision determines the cost share of each user. The cost share "... will influence the location of equilibrium via income effect feedbacks on demand."[29] The solution is now determinate since the utility manager has selected one of an infinite number of possible price offers which could be made. However, its selection entails a distributive judgment on the part of the utility manager.

The type of model described by Buchanan has particular significance to the practices of municipally owned water utilities. Buchanan's conclusions rest on the assumption of first degree price discrimination in infra-marginal units. In municipalities, there is considerable variation of rates between some types of users. The rate variations can be used for taxation or for community development. The community leaders make the selection of the price offers, and, hence, the distributive judgment about the basis for rates. Buchanan states that the efficiency and

---

[28] Ibid., p. 466.

[29] Ibid., p. 466.

distributive aspects cannot be separated. In Buchanan's words:

> Choosing a specific distribution of the total cost among separate-period demanders will, of course, determine a specific allocation of resources and a specific set of marginal prices that must be obtained if Pareto efficiency is to be achieved. On the other hand, within certain limits, choosing a specific investment in system capacity will determine the distribution of total costs along with the set of marginal prices that must be present if the Pareto conditions are to be satisfied. In either case, the limits of economic analysis are reached sooner than...the Steiner...analysis implies.[30]

The position taken by Buchanan further reinforces the view that additional economic analysis is needed to guide municipal water utility managers to make efficient decisions with respect to plant size and rate structures. There is an even greater need for economic evaluation of the various distributive judgments these same managers make, and to which Buchanan explicitly draws attention. Previous economic models have failed to provide this evaluation.

The Underlying Assumptions

The various solutions to the peak load problem all contain assumptions which either enhance or limit their applicability to municipal water utility problems. In other respects, the authors have failed to make explicit the nature of their assumptions. It becomes important to examine these assumptions,since the assumptions affect the validity of

---

[30]Buchanan, "Peak Loads and Efficient Pricing: Comment," p. 471. Buchanan refers the reader to: R.H. Strotz, "Two Propositions Related to Public Goods," Review of Economics and Statistics, XL (November, 1958), pp. 329-31.

the model in terms of its applicability to the problems faced by municipal water utilities.

Joint Supply

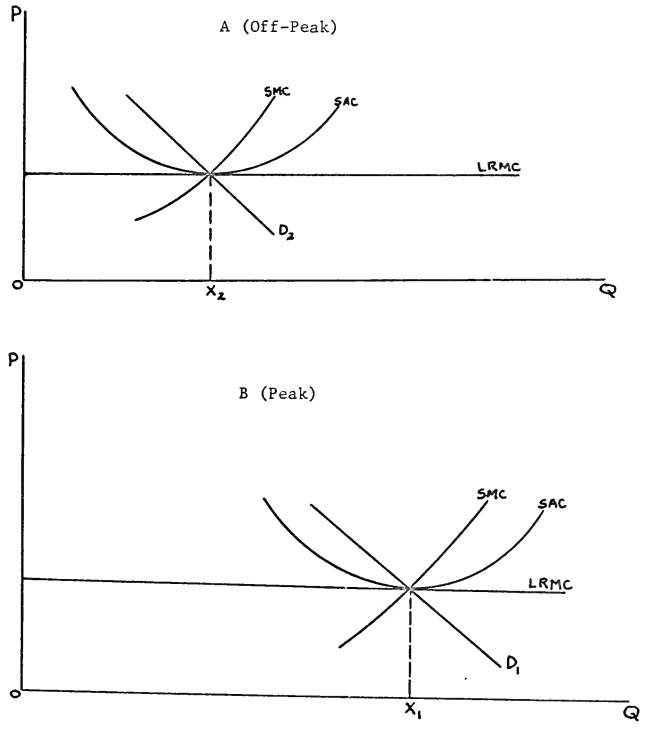
The first solution to the peak load problem was the Steiner solution which included the vertical summation of the peak and off-peak demands to solve for optimal capacity and the optimal rate structure. The use of vertical summation--a tool used also by Williamson and Buchanan, but not Hirshleifer--recognizes the joint nature of the peak and off-peak periods in which the capacity used to provide output in one period is the same capacity used to provide output in the other period. In using vertical summation, Steiner employs the straightforward summation which is valid when the time periods are of equal duration, whereas Williamson, recognizing that the time periods are not of equal duration, introduces a weighted summation to allow for the difference. Williamson weights each demand by the proportion of the time it represents of the entire cycle.

A significant aspect of these solutions is their recognition of the joint supply problem. However, the authors leave their readers with the impression that the output produced during the off-peak period is produced jointly with the output produced during the peak period, but such is not the true joint nature of the problem. The jointness is in respect to time and capacity. The jointness arises because of the firm's inability to vary neatly the size of its plant between the peak and off-peak periods.

If capacity could be changed between the two periods, plant size would vary in accordance with the illustrations in Figures $XIII_A$ and $XIII_B$. The off-peak case, shown in Figure $XIII_A$, yields a capacity

Figure XIII

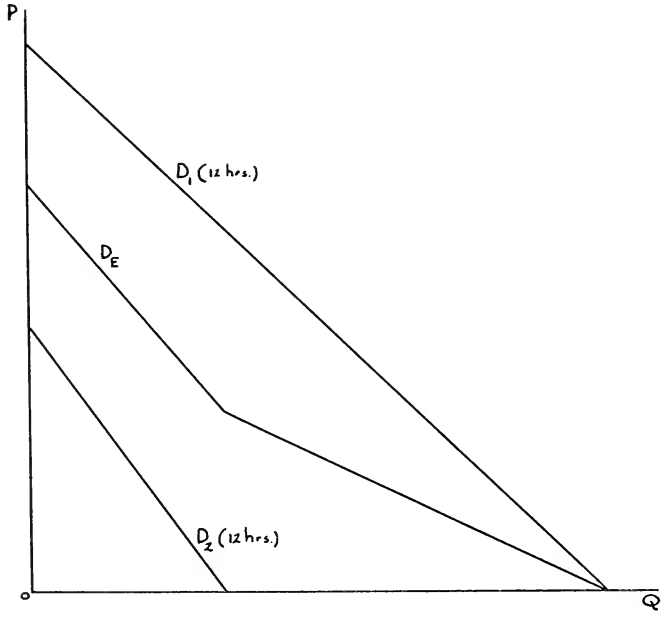
The Optimal Solution in the Absence of Time Jointness



sufficient to produce output OX_2. As the peak demand becomes the effective demand, the utility expands its capacity to produce output OX_1 shown in Figure $XIII_B$. As the peak demand declines, and the off-peak demand becomes the effective demand, plant capacity is reduced back to that shown in Figure $XIII_A$. The inability to vary the amount of capacity in this manner over the short-run creates the joint supply

Figure XIV

The Vertical Summation
of Intra-Cycle Demands



condition since the firm is faced with two demands in the same time horizon and is unable to adjust its capacity neatly to each of these demands. Consequently, the capacities between the two subperiods are provided in fixed proportions to each other. The proper price and output decisions are, therefore, based upon a vertical summation of the two subperiod demands. The vertical sum shows the average price the firm receives over the entire time horizon under consideration. The weighted summation, shown in Figure XIV as the demand function D_E under the assumption of periods of equal duration, is applicable to a water utility since this type of firm must produce to meet seasonal, daily, and hourly demands which can exceed the average demand by as much as 1500 per cent. The inability to adjust plant capacity to these demands generates a true case of time jointness in the operation of a water utility.[31]

Administrative Problems: Price Stability and Technological Deficiencies

There are several problems a water rate analyst may encounter in attempts to administer prices based on an efficiency criterion. One of these problems entails the possibility of continually altering prices

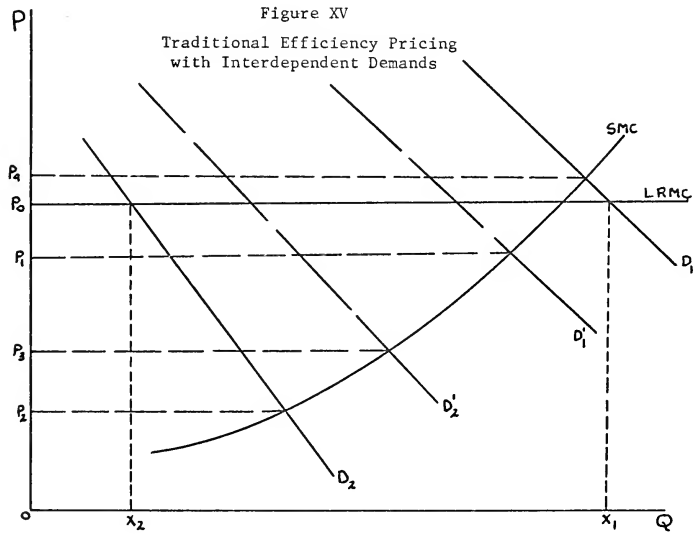
---

[31]Storage might be a feasible alternative for meeting the hourly peak demand. Storage tanks and reservoirs may be filled during the off-peak period to aid in meeting the heavy demand occurring during the peak period. The use of storage facilities enables the utility to partially adjust capacity between the peak and off-peak periods. However, storage is not a perfect substitute for production and pumping capacity. Therefore, the existence of storage will modify but will not drastically change the conclusions.

when the peak and off-peak demands are interdependent, and price differentials cause a shifting peak. If the utility charges a single price for providing both peak and off-peak water service, the rate structure clearly fails to achieve Pareto optimality. Figure XV demonstrates that an attempt to use efficiency pricing, when demands are interdependent, might

Figure XV

Traditional Efficiency Pricing
with Interdependent Demands



conceivably lead to difficulties in making the price adjustment. In Figure XV, the original price is OP_0 (the single price), the quantity taken during the off-peak is OX_2, and the quantity taken during the peak is OX_1. At this level, price is less than the peak short-run marginal

cost and greater than the off-peak short-run marginal cost, and the Paretian criterion is not satisfied. If each period's price is to equal short-run marginal cost, the off-peak price should be lowered to OP_2 and the peak price increased to OP_4. However, as the off-peak price is lowered, and the peak price is increased, the users will shift their demand from the peak hours to take advantage of the lower off-peak rate. This shifting causes the peak demand (D_1) to move downward and to the left, and the off-peak demand (D_2) upward and to the right--assume to D_1' and D_2' respectively. To maintain prices equal to short-run marginal cost, the peak price must be lowered to OP_1 and the off-peak price increased to OP_3. However, as the differential between the peak and off-peak prices is reduced, the demands will move back toward D_1 and D_2 respectively, as the users now shift part of their demand away from the off-peak hours and back to the peak hours. This movement requires raising the peak price and lowering the off-peak price to keep prices equal to short-run marginal cost. Gross adjustments of this sort might lead to a fruitless back and forth adjustment of prices. However, small adjustments, taken one at a time, may lead to a converging cobweb.

The problem of interdependent demands combined with shifting peaks becomes involved in a highly theoretical analysis of dynamic economic equilibrium entailing such factors as the nature of price stability, the existence of a single equilibrium versus multiple equilibria, and the problem of price determinacy. The area should be recognized as a problem area worthy of further investigation, but the investigation is beyond the immediate scope of this study. In any case, interdependence complicates the pricing problem, especially in the case of shifting peaks.

Another administrative problem associated with prices based upon rigorous efficiency criteria is the technological deficiencies present in today's water meters. Assume the utility is able to vary the peak and off-peak prices to reflect cost differences in satisfying the hourly water demands without the complexities of interdependent demands and shifting peaks. To arrive at a proper pricing policy, water use must be measured on an hourly basis. However, at the present time, the metering equipment necessary to accomplish this task is far from being perfected to the point which could make this a relevant alternative. The present cost of such metering equipment is so high that any gain through a rigorous price policy probably would be offset by increases in costs which would accompany the installation of the metering equipment. Thus, flat monthly rates, which seem not to conform to the marginal cost criterion, indeed, may represent the most efficient pricing arrangement. Efficiency, in any real sense, must encompass all costs including administrative costs.

These administrative problems associated with a price policy rigorously tied to marginal cost pricing may require some alternative form of pricing such as the use of fixed charges. These alternatives are developed in Chapter V.

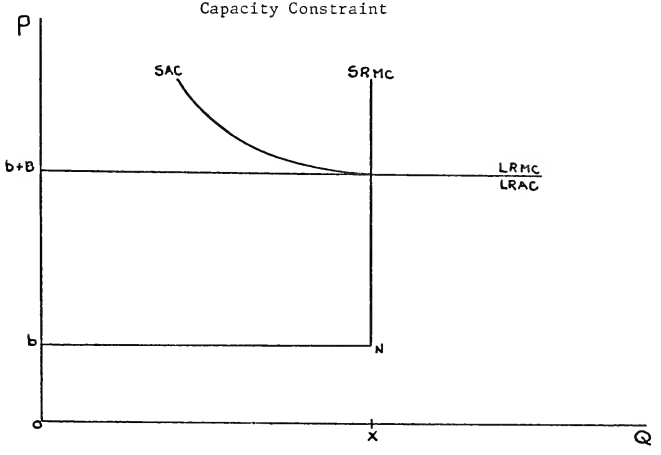
Cost Functions

All the models assume constant returns to scale as the utility expands, an assumption retained, in part, in the remainder of this study. There is, however, a definitional point which requires clarification. There appears to be some ambiguity arising from the use of various interpretations of short-run marginal costs when dealing with a straight line discontinuous

marginal cost function. To show a capacity constraint, the models assume a cost relationship such as that illustrated in Figure XVI. At output OX, the firm is unable to produce additional output, and the short-run marginal cost function becomes a vertical line. The implication is that output OX defines the point where the short-run marginal cost function becomes infinite. To achieve economic efficiency, the models

Figure XVI

Cost Functions with a Capacity Constraint



call for prices equal to the short-run marginal costs.[32] The horizontal

[32]One exception is Davidson whose solution equates price with long-run marginal cost. Davidson, Price Discrimination in Selling Gas and Electricity.

segment (bN) of the short-run marginal cost function is the relevant segment from the standpoint of production costs. This part represents the short-run marginal operating (variable) costs. The vertical segment beginning at point N, although labeled short-run marginal cost, is not an operating cost, and, thus, not short-run marginal cost. The traditional neo-classical concept of the marginal cost of production is the change in total costs associated with a change in output.[33] Once output OX is reached, the firm cannot increase output without adding to its capacity. All inputs become variable, and an attempt to increase output without adding to capacity would be non-economic behavior since it cannot be done. For additional output beyond OX, the relevant cost function becomes the long-run marginal cost function and not short-run marginal costs. To refer to the vertical segment as either opportunity costs or social costs is to vary the definition of marginal cost along the same cost function, a practice which easily leads to confusion.

The models tend to employ three different concepts of short-run marginal cost. The horizontal segment represents short-run operating costs. When the firm comes up against the capacity constraint, Hirshleifer changes the definition to one of marginal opportunity cost while Hotelling uses the marginal social cost of congestion concept. Steiner, Williamson, and Buchanan do not resort to the marginal opportunity or social costs concept since they are treating the problem as a joint cost problem and use the

---

[33]For the traditional concept of marginal costs see: George J. Stigler, The Theory of Price (rev. ed.: New York: The Macmillan Co., 1962), pp. 96-97.

vertical summation of demands.[34] It can be seen that several concepts of marginal cost are used in the models, and it is important that these different concepts be made explicit, as they make an important distinction between the approaches of the various models.

The Equity Implications

The basic solutions to the peak load problem presented by the economic models have been those that require equating price with marginal costs. All price differences are explained by costs, with any differences in costs reflected in the prices. With the exception of Buchanan, all distributive judgments have been assumed away either implicitly or explicitly. The solutions are based upon the criteria of economic efficiency, but these efficiency solutions have distributional implications.

The Peak Problem

All the solutions advocate that the peak users pay all the capacity costs when the off-peak users fail to utilize all the capacity. The conclusion is based upon a view that only the peak user imposes capacity

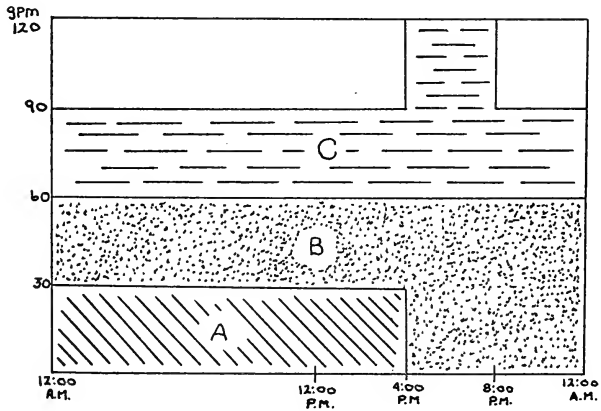
---

[34]It was this point that led Steiner to the conclusion that prices set along the vertical short-run marginal cost function are discriminatory. Steiner, "Peak Loads and Efficient Pricing," p. 590. The significant point Steiner made was that, although the prices are discriminatory since they are based upon demand and not cost, the solution is consistent with Pareto optimality since the sum of the prices is equal to long-run marginal cost. The other writers either ignored the point made by Steiner or, as did Hirshleifer, redefined the vertical segment in terms of some other form of costs. Hirshleifer, "Peak Loads and Efficient Pricing: Comment," pp. 458-59.

costs.[35] A question of equity arises in the case of the off-peak user, who appears to go scot-free. Figure XVII shows the hourly load chart for a water utility serving three users (A, B, and C). User A takes 30 gallons per minute (gpm) every hour between 12:00 A.M. and 4:00 P.M. User B takes 30 gpm during each hour between 12:00 A.M. and 4:00 P.M. plus 60 gpm for each hour from 4:00 P.M. to 12:00 A.M. User C takes 30 gpm for each hour from 8:00 P.M. to 4:00 P.M. and 60 gpm during each hour from 4:00 P.M. to 8:00 P.M. Using the capacity cost allocation advocated by

Figure XVII

An Hourly Load Chart for Three Users



[35] A conclusion reached by: Steiner, "Peak Loads and Efficient Pricing," pp. 585-610; Hirshleifer, "Peak Loads and Efficient Pricing: Comment," pp. 451-62; Williamson, "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," pp. 810-27; Davidson, Price Discrimination in Selling Gas and Electricity. Williamson and Davidson are the most explicit on this point.

the theoretical models, total capacity costs are recovered during the four-hour period, 4:00-8:00 P.M., and, in this example, only users B and C pay capacity costs while user A pays only the variable costs of his service and makes no payment toward the capacity costs. The conclusion is that, if user A reduces his consumption, there is no change in the capacity costs of the system. User A, however, uses part of the capacity during the off-peak period, and, if users B and C were to reduce their consumption (or cease their consumption), all capacity costs would not be eliminated. Part of the capacity would have to be retained to continue serving user A. If the peak user pays all the capacity costs, user A becomes a free rider. The general conclusion of the models is based upon the assumption of jointness,[36] and the distribution of the cost shares is based upon the relative strength of the various demands. However, to let some users acquire water service without making a contribution to the capacity costs might be deemed "inequitable."

A second distributional problem arises when the peak hours or months do not contain a level load. Pushing the models' cost distribution to their limits, all capacity costs would be recovered in the price of the water used during one second or one minute of the year. The problem is demonstrated in Figure XVIII. Given this load distribution, all the

---

[36]The exceptions to this generality are the Hirshleifer and Davidson models. Hirshleifer treats the problem as one involving opportunity cost. Davidson's analysis appears to be erroneous and uses average cost pricing rather than marginal analysis. Hotelling treats the problem as one involving social congestion costs.

capacity costs are recovered in prices charged for water used at precisely 6:00 P.M. Putting all the capacity costs into the charge for water used at one moment in time makes the charge exhorbitant. The high price, along with the high costs of measuring water use during each minute in the year, makes this alternative impractical as well as one that is possibly inequi-

Figure XVIII

An Hourly Load Chart

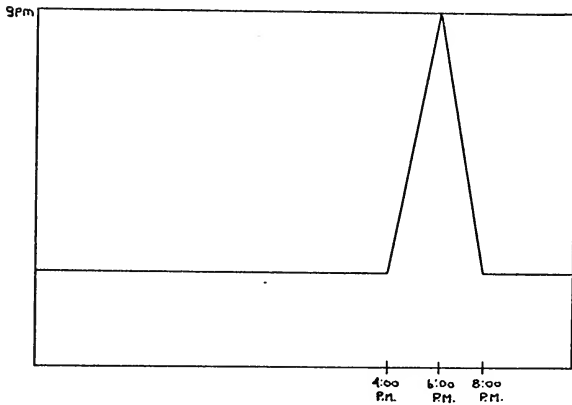


table and inefficient (since costs of administration must be considered in any efficiency situation). The most feasible alternative is to average the water quantities used over the four-hour period and spread the capacity costs over the entire peak period. If this alternative is used, there is a series of redistributive effects. The water user at 6:00 P.M. pays less capacity costs than he does under the pure economic solution.

An individual user taking water at 5:45 and 6:15 P.M., but not at 6:00 P.M., now pays part of the capacity costs, but makes no contribution to capacity, according to the economic solutions. The water users at 6:00 P.M. receive their service at below cost since they pay less than the capacity costs they impose upon the system.

Costs Versus Ability to Pay

The requirement that users pay in accordance to the marginal costs they impose becomes tangled in a distributional problem. The use of a price policy consistent with marginal costs may be in conflict with the ability to pay on the part of the user. People using water early in the morning and during the early evening hours are the users who must bear the full capacity costs. Users who have to use water during the peak hours because of their work habits, and, hence, must pay the capacity costs, might not be the same users who have the ability to pay the capacity costs; they may be unduly burdened. The imposition of the greatest part of the capacity costs on these users might result in a distributional effect the community considers to be undesirable.

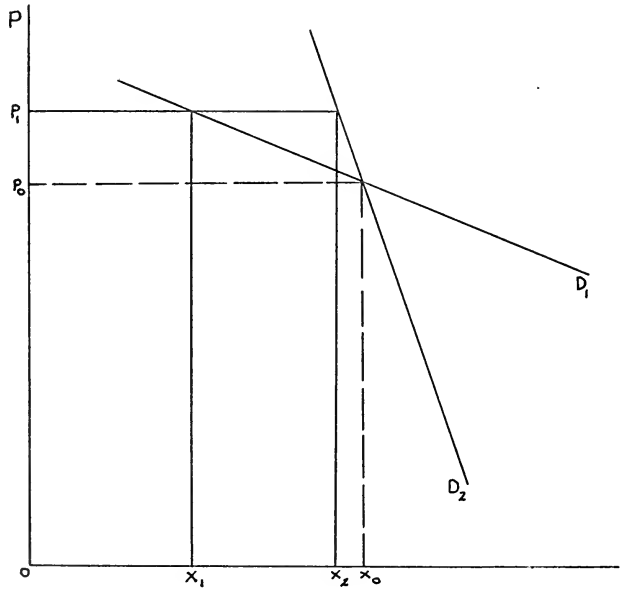
Elasticity of Demand and Resource Impact

It is significant to note that elasticity of demand plays an important role in determining the effects of price changes on the total level of resource use. The impact on resource use varies inversely with elasticity of demand. This impact is illustrated in Figure XIX where a change in price from OP_o to OP_1 is shown to have differential effects on the quantity produced and the size of plant. If prices are to be manipulated, resource impacts are minimized when the manipulations relate to

the inelastic demand. If D_1 (Figure XIX) represents the demand for electricity, while D_2 represents the demand for water, a community which controls the prices of both services may elect to derive general revenue from its water operations so as to minimize the negative adjustment in resource use. It is even possible that the water utility may subsidize

Figure XIX

Elasticity and the Impact of Price



the electric utility, subsequently leading to a net increase in the level of resource use in the community. The same general relationships, of course, apply in the case of the customer classifications of each utility. Thus, price discrimination may be structured so as to increase resource use in the local economy, or to raise general revenue with a minimum decrease in resource use.

Since the demand for water probably is more inelastic than the demand for electricity, one would predict greater reliance on water utilities for general funds than on municipally owned electric plants. Data for such comparisons are not available. However, tradition generally favors profitable electricity operations and frowns on the sale of water under a value-of-service principle. These practices probably are explained better in terms of historical patterns in the institutional development of the two types of utilities.

It should also be noted that although resource impacts are small, in the presence of demand inelasticity, income effects may be large. Thus, a serious equity problem arises when the utility takes the easy way out by raising the price of the commodity which is in inelastic demand. If those users who have inelastic demands also happen to have low incomes, a policy justified in terms of minimizing resource effects may lead to unacceptable distributional consequences.

Conclusions

This chapter has attempted to demonstrate that the economic models have made significant inroads into the problems of pricing and investment in the cases of firms faced with a peak load problem. Some of the solutions have combined marginal cost pricing with the vertical summation of demand functions to handle the problem of time jointness between the peak and off-peak demands. However, these theoretical models have been weak on some points such as the use of discontinuous marginal cost functions, which, in the case of Hotelling and Hirshleifer, leads to a shifting back and forth between operating costs, opportunity costs, and congestion costs along the same cost function. The strict application of capacity costs to the peak user leads to rate structures which can either be too high to be practical, or else permit one user to acquire water service at prices below marginal cost while other users pay rates in excess of marginal costs. It is significant to note that the Davidson solution, which was intended to be a marginal cost pricing solution, appears to be a full cost distribution resulting in average cost pricing.

The basic problem in the models is the elimination of the distributional problem. Although the rates suggested may conform to efficiency criteria, the models have developed rate structures which might have undesirable distributive effects. The models do not take into consideration the distributional aspects of their efficiency solutions. The impact becomes significant because a conflict between efficiency pricing and distributional equity requires trade-offs. The failure to evaluate

the distributional impacts requires more analysis to make the theoretical models directly applicable to the problems faced by a municipally owned water utility.

CHAPTER IV

PRACTICAL DESIGN OF WATER RATES

Introduction

This chapter describes, highlights, and evaluates those concepts and techniques that are currently employed in the practical design of water rates. It assesses the implications of current practice for the efficient utilization and development of water facilities. A development of models which attempt to describe these water utility practices is presented in Chapter V.

The major criteria employed in the practical design of utility rates are (a) the cost-of-service, (b) the value-of-service, and (c) competition.[1] Properly interpreted, these criteria represent an effort to determine rates in some relation to the forces of supply and demand. Since there is no near substitute for water, at least within any realistic range of prices, it is generally conceded that water supply must be provided under monopoly conditions in order to avoid duplication and to capture economies of scale. Therefore, the influence of competition on demand elasticity is not a significant rate-making factor. However, competition may play a significant role in determining prices at the extensive geographical margin in the case of disputed or overlapping territories

[1]For an excellent discussion of these and other rate-making criteria see: Eli Clemens, Economics and Public Utilities (New York: Appleton-Century, 1960), pp. 247-369; D. Philip Locklin, Economics of Transportation (6th ed.; Homewood: Richard D. Irwin, Inc., 1966), pp. 130-57.

and, in some instances, where wells provide an alternative source of supply. Notwithstanding these possibilities, competition is not a standard factor in deliberations concerning the design of water rates. Similarly, value-of-service seems to play a very limited role in the design of water rates although it may be very important in actual practice as price discrimination takes place. High rates charged on the basis of value-of-service violate the tradition that the price of water should not reach a monopoly level. Low rates to certain users are usually justified in terms of an "out-of-pocket" cost criterion.

Regardless of the rationale by which competition and value-of-service are excluded, the literature on water rate design is almost entirely devoted to a "fully allocated cost-of-service" criterion. In this respect, the criteria for water rate design differ from those used in the case of most utility services. It can also be argued that the models developed by utility practitioners fail to satisfy their own criteria for equity and efficiency.

The Cost Allocation Technique

The Bias Towards Costs

The bias towards full cost-of-service criteria in water utility ratemaking is related to a tradition which views cost as a criterion of both efficiency and equity. According to the literature, a rate structure is "equitable" if each user pays in accordance with the costs assignable to

him.[2]

The total revenue requirements imposed upon a municipally owned utility are oriented toward the recovery of all costs. Indeed, the major criterion of efficiency and success relates to the ability of the rate structure to recover costs.[3] To be financially viable and "efficient" (according to current thinking), the municipally owned utility should recover operating and maintenance costs, interest and amortization of the public investment, reserves to provide distribution mains, meters and meter servicing, and a payment to the community's general fund.[4]

The general fund contribution is usually viewed as a cost payment to the city in lieu of taxes. Obviously, the revenue requirements are entirely cost oriented. Demand factors are simply not discussed in the

---

[2]William L. Patterson, "Practical Water Rate Determination," Journal of the American Water Works Association, LIV (August, 1962), pp. 905-6; Louis R. Howson, "Review of Ratemaking Theories," Journal of the American Water Works Association, LVIII (July, 1966), p. 855; Charles W. Keller, "Design of Water Rates," Journal of the American Water Works Association, LVIII (March, 1966), p. 296; Jerome W. Milliman, "The New Price Policies for Municipal Water Service," Journal of the American Water Works Association, LVI (February, 1964), p. 127; Staff Report, "The Water Utility Industry in the United States," Journal of the American Water Works Association, LVIII (July, 1966), p. 773.

[3]E.D. Bonine, "Making a Water Utility Solvent," Journal of the American Water Works Association, XLV (May, 1953), p. 457.

[4]This breakdown appears to be the generally accepted criteria for a utility's revenue requirements. See: Patterson, "Practical Water Rate Determination," p. 904; Howson, "Review of Ratemaking Theories," p. 850.

rate-making literature.[5]

The bias towards the supply, or cost-of-service, side is also reflected in the various customer and rate classifications used by municipal water utilities although these classifications are not of great importance for rate-making purposes.[6]

Water utilities generally appear to recognize three classes of customers: (1) residential; (2) commercial; and (3) industrial. To the extent that these classes are considered for rate purposes, it is with respect to the special costs each class imposes. Generally, the same block-rate schedule applies to all users, although this is not universal. Sometimes, distinctions are made between classes on the basis of the meter deposit--the larger the meter, the larger the deposit. Since meter costs are a special cost and not related to the costs generated

---

[5]See: Charles E. Howe and F.P. Linaweaver, Jr., The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structures (Washington: Resources for the Future, Inc., 1967), p. 13. These writers go so far as to charge that utility managers pass over demand altogether with an implicit assumption that demand is completely inelastic at an infinite quantity. These writers point out that "... the demands the system is to be designed to meet depend upon the price charged; the price charged must be related to the costs of the system if the utility is to be economically efficient and financially viable; but costs are determined by design which depends upon demand." The writers maintain that this circle is ignored in the engineering and utility management literature.

[6]For elaboration of this point see: Orville P. Devel, "Water Utility Rate Making," Journal of the American Water Works Association, LVIII (July, 1966), p. 846. Devel explains that one exception is the use of fire protection charges. These charges incorporate some consideration of the property value for which protection is provided. At the same time, Devel makes the point that "...no one rate making procedure for private fire protection service exists that is clearly defined, or has wide acceptance."

by water use, the justification for the distinction between meter sizes is based upon these special costs. A larger meter requires greater maintenance and entails greater installation and depreciation expense. Moreover, it is associated with a larger capacity requirement.

There are four significant variables that affect the costs of providing water service. The variables are: (1) the number of customers the utility serves; (2) the total quantity of water the utility produces; (3) the rate at which water is produced; and (4) the geographical characteristics of the area and the population distribution within the area. These variables are evaluated by water rate analysts in terms of their effect upon the design of the water system, i.e., costs. There is very little analysis in terms of the impact that rate structures have on the quantity of water demanded and, hence, on the system design. The usual analysis appears to be concerned only with the impact these variables have upon the utility's costs.

The cost allocation used for rate-making purposes is based upon a threefold classification of total costs: (A) customer costs, (B) base water costs, and (C) extra-capacity costs.[7]

A. Customer costs are the costs of servicing an individual customer without regard to the quantity of water used. These costs include the

---

[7]This breakdown is the one generally accepted in the field. See: Keller, "Design of Water Rates," p. 294; Patterson, "Practical Water Rate Determination," p. 909; Milliman, "New Price Policies for Municipal Water Service," p. 130; Jack Hirshleifer, James C. DeHaven, and J. W. Milliman, Water Supply (Chicago: University of Chicago Press, 1960), p. 98.

costs of meter installation and maintenance, meter reading, accounting, billing, and collections. The customer costs are directly related to the number of customers the utility serves.

B. Base water costs are the costs of providing a given supply under optimal load conditions, i.e., at a 100 per cent load factor. These are the costs if the utility's demand is such that each customer uses 1/365 of his annual consumption each day, and 1/24 of each day's use every hour. Base water costs include production costs, distribution expenses, interest, and amortization of the utility's investment. The base water costs vary directly with the total quantity of water used and the rate of production.

C. Extra-capacity costs are the costs in excess of base water costs which arise when demand is such that water cannot be provided according to a 100 per cent load factor. The actual demand for water at any given time is usually greater or less than the average daily demand.[8] Water use has peaks and troughs depending upon the month of

---

[8] In a study of Baltimore County made by J.B. Wolff, it was found that: (1) well established older neighborhoods had a peak hourly demand 500 to 600 per cent of the average daily demand; (2) for newer neighborhoods with average size lots of 1/4 to 1/2 acres, the ratio of peak hour demand to average daily demand was 900 per cent; and (3) in both newer and older neighborhoods with large lots of 1/2 to 3 acres, the ratio was 1500 per cent. J.B. Wolff, "Forecasting Residential Requirements," Journal of the American Water Works Association, XLIX (March, 1957), p. 233. Also see: Devel, "Water Utility Rate Making," p. 847; William L. Patterson, "Comparison of Elements Affecting Rates in Water and Other Utilities," Journal of the American Water Works Association, LVII (May, 1965), p. 556.

the year and the hour of the day. Residential water use is higher in the early morning and early evening than it is during the other times of the day, and is higher in summer months than during the winter months. These variations in water use require a large reserve capacity to meet the peak demands and to avoid interrupting service. The reserve capacity remains idle during the off-peak or slack months and hours. The need for additional capacity is the basis for the extra-capacity expenses resulting from the extra capacity beyond that required for a 100 per cent load factor.[9] No costs for pumping power or chemicals are included since these costs are included in the base water costs.[10]

The Patterson Allocation

The method by which the customer and cost classifications are used to determine a water rate structure is presented by a model developed and illustrated by William L. Patterson.[11] According to Patterson, "... to be equitable, the application of the rate schedule should reflect the effect of peak demand for service, as well as the total water use and other costs for servicing customers in all various classifications."[12]

---

[9]Patterson, "Practical Water Rate Determination," p. 907.

[10]Ibid., p. 907.

[11]Ibid., pp. 904-12.

[12]Ibid., p. 906.

The first problem Patterson undertakes is the allocation of total costs among the three basic cost categories. In his example, maximum demand is 300 per cent of the average demand. Since average demand is 1/3 of the peak demand, 33 per cent of the production and distribution costs are categorized as base water costs. Storage tanks required to meet maximum demand are extra-capacity costs. Customer meter expenses and services are assigned directly to customer costs. Once the total costs are allocated, the total base water costs and extra-capacity costs are divided by the number of 1,000 gallon units produced and the result is the cost per 1,000 gallons. The customer costs are divided by the number of equivalent 5/8" meters to determine the per-unit customer costs. In Patterson's model, base water costs are $0.087 per 1,000 gallons annually, extra-capacity costs are $21.21 per 1,000 gallons annually, and customer costs are $10.21 per equivalent 5/8" meter annually.

The next problem is to allocate the costs in each category to each of the customer classes. In his model, Patterson bases extra-capacity cost requirements on the maximum class demands of 400 per cent for residential, 250 per cent for commercial, 200 per cent for industrial, and 150 per cent for "special." The figures are percentages of the average rate of use for each customer class. Based upon these ratios, the residential users are allocated 64 per cent of the base water costs and 81 per cent of the extra-capacity costs. The commercial user pays 20 per cent of the base water costs and 13 per cent of the extra-capacity costs. The industrial user pays 12 per cent of the base water costs and

5 per cent of the extra-capacity costs. The special user is assessed 4 per cent of the base water costs and 1 per cent of the extra-capacity costs. Customer costs are assigned on the basis of the number of equivalent 5/8" meters in each customer classification. Each cost component is summed to arrive at the total costs assigned to each user class. The final allocation is 78 per cent to the residential users, 15 per cent to the commercial users, 6 per cent to the industrial users, and 1 per cent to the special users.

The final step is to develop the rate structure. Patterson sets up his model so that revenues are less than costs in order to demonstrate how rates in each class should be increased to make revenues equal to costs and to achieve "equitable" rates. The total costs and revenues in each customer class are compared to determine the percentage rate increase that is required to make total revenue from each user class equal to the total costs assigned to each user class. In the Patterson model, the average rate increase is 30 per cent, but the individual customer class increases range from 27 per cent for the residential rate to 50 per cent for the industrial rate. Once the percentage increases are determined, Patterson then demonstrates how the per-unit costs are converted into rates. He uses a monthly minimum charge with a 2,000 gallon allowance combined with a declining block-rate schedule and, then, derives the minimum charge, the rate on the first 1,000 gallons above the service charge, and the rate on the last 1,000 gallons in the last block of the schedule. Patterson applies the per-unit figures derived from the allocation of total costs between customer costs, base water costs, and

extra-capacity costs. The result is a minimum charge of $1.66 which includes the first 2,000 gallons, a rate of $0.264 for the first 1,000 gallons in excess of the minimum, and a rate of $0.117 on the last 1,000 gallons in the rate structure.[13]

The model presented by Patterson is typical of the method advocated for rate determination by water rate consultants. The final rate structure is a service charge plus a declining block schedule which is applicable to users in each class. Separate schedules are recommended for each class of user (Patterson uses only the total for simplicity).

Types of Charges

There are at least eight types of charges used by water utilities to collect revenue from customers. Although no utility uses all eight charges, a combination of two or three of the charges is usual. The charges are: (1) fire protection charges, (2) front-foot assessments, (3) tap-on charges, (4) flat-rate charges, (5) service charges, (6) minimum charges, (7) declining block-rate commodity charges, and (8)

---

[13]The monthly service charge is determined by the sum of the base water costs, extra-capacity costs, and the customer costs. The base water component is determined on the basis of 8.7 cents per 1,000 gallons, which, with a minimum allowance of 2,000 gallons, makes the cost 17.4 cents. The extra-capacity costs are $21.21 per month, and, on the basis of 2,000 gallons per month, converts to 35 cents. The customer costs are $1.14. The monthly service charge is 17.4 cents plus 35 cents plus $1.14 which equals $1.66. The rate for the first 1,000 gallons above the minimum is made up of base water costs, which are 8.7 cents, and extra-capacity costs of 17.7 cents. The rate for the last 1,000 gallons in the last block is 11.7 cents--base water costs of 8.7 cents and extra-capacity costs of 3 cents. The latter is obtained by converting the $21.21 into a 1,000 gallon charge. Ibid., p. 910.

demand rate charges. These charges may be classified as either fixed charges (charges 1-6) or variable charges (charges 7-8) depending on their variation with respect to water use.

Fixed Charges

Of the fixed charges listed above, two are oriented toward both costs and benefits, and the other four are entirely cost oriented. Front-foot assessments and fire protection charges are primarily benefit charges which are related either directly or indirectly to property value.[14] These two charges, however, are the least significant in terms of actual use.

The fire protection charge is usually based upon the property value along with the number of hydrants located at the site with the emphasis placed on the latter. The combination of property value and the number of hydrants into the charge gives the charge both value-of-service and cost-of-service aspects. However, there appears to be very little analysis of the proper allocation of costs between fire protection charges and the charges on normal water use which would seem to indicate a pragmatic approach to the allocation.

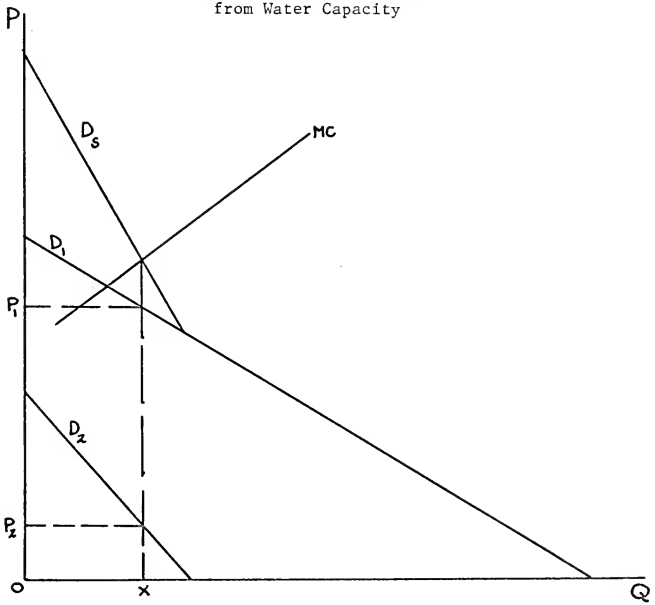
Vertical summation appears to provide a method of determining the allocation of costs between fire charges and normal use charges. Figure XX shows the benefits of normal water use (D_1) derived from water capacity

---

[14]Louis E. Ayers, "Methods of Financing Water Utilities in Michigan," Journal of the American Water Works Association, LI (January, 1959), pp. 10-11.

Figure XX

The Combined Benefits Derived
from Water Capacity



(measured along the horizontal axis), and the benefits derived from this capacity for fire protection (D_2). The summed demand (D_S) is the combined benefits derived from the capacity, and MC is the marginal cost of supplying capacity. The proper amount of capacity is OX, where the combined benefits are equal to the costs of capacity. The cost allocation between normal water use and fire protection is such that price OP_1 is the rate charged for normal water use, and price OP_2 is the rate charged

for fire protection. The solution combines the value of service with the costs of service to reach a Pareto optimum. The use of fire protection charges which are directly related to property values seem to have a regressive effect with respect to income. Income is redistributed from low income recipients to high income recipients.[15]

The use of front-foot assessments is only vaguely related to property value, but it can result in an efficient cost allocation when some costs are not related to water use. The underlying principle is that everyone pays an equal price for each front-foot of property. Consequently, a family of eight with 100 front-feet pays the same price for their water as a family of two with the same front-footage.[16] However, when a water utility puts its water mains along the street, the cost of the main is not related to the amount of water use, but to the number of front-feet of property. In these cases, property owners pay a price equal to the special costs of providing service. Hence, there may be little redistributive effect on income, as each user pays a price equal to his cost of

---

[15]This conclusion is based on the assumption that all property taxes are passed on in the case of rental property. Since a greater percentage of low income classes rents relative to the percentage of the high income classes that rents, the overall effect is a regressive incidence of any rates tied to property values. For an excellent study of property taxes see: Maryland, Guidelines for Improving Maryland's Fiscal Structure (Interim Report: January, 1965), p. 109.

[16]If all water costs are recovered in this manner, and some costs are related to water use, an inefficient cost allocation might result. The family of two will in all probability pay a price which exceeds its cost of service while the family of eight will probably pay a price less than its costs of service. Consequently, income will be redistributed from the small family to the large family as the latter receives an "unwarranted rate subsidy."

service. The balance of the water utility's costs can be recovered through another type of charge.

The tap-on charge is a "once and for all" charge collected at the time the water line is connected to the street main. The charge is designed to recover the connection costs which are assignable to any given user, and usually varies with the size of the connection.

The flat-rate charge is applied in one of two ways. It is applied as a flat rate per month when water use is not metered. On the other hand, if water consumption is metered, the charge is normally applied as a flat rate per 1,000 gallons per month. In the first instance (flat-rate), the user pays a fixed amount each month regardless of the amount of water consumed. The rate is determined by dividing total costs of service by the total number of customers. The annual cost is divided by twelve to arrive at the monthly charge, which is then set high enough to recover the total costs of water service plus any contribution the utility is expected to make to the city's general revenue.

When water use is metered, the user pays a fixed amount for each 1,000 gallons of water used per month. The rate is determined by dividing the total amount of water produced into 1,000 gallon units and then dividing this amount into total costs. This result is divided by twelve to arrive at a monthly charge per 1,000 gallons.

The service charge and the minimum charge are similar, and are often used interchangeably. The primary distinction is that the service charge does not include a minimum amount of water and is a true minimum. Customer objections to such charges arose during the Thirties when many persons objected to a payment when no water was used. Utilities reacted by

raising the service charge and allowing a minimum gallonage "free of charge." Consequently, the minimum charge includes a certain amount of water which ranges from 1,000 to 6,000 gallons per month.

The service charge and the minimum charge are designed to be used in conjunction with a declining block-rate commodity charge. They are designed to recover the customer costs, part of the base water costs, and part of the extra-capacity costs.

Variable Charges

Two basic variable charges are presently used by water utilities. These two charges are the declining block-rate commodity charge and the demand commodity charge. These charges are designed to recapture costs in accordance with the quantity of water taken by each user, and the rate of the user's water consumption. As water use increases, the utility is able to supply water to the customer at decreasing costs (assuming, of course, that it has available capacity). These charges include both the balance of the extra-capacity costs and those base water costs not included in the basic water service charge.

Of all the possible charges, the most widely used combination appears to be a minimum charge accompanied with the declining block-rate charge. The reason for the popularity of this combination seems to be based upon the belief that users are more familiar with a minimum charge which allows a certain gallonage. Another reason is that utility managers seem to consider this arrangement to be the most "equitable" combination. The use of a declining block-rate schedule is justified on cost grounds. The amount of water used by the large users enables the water plant to realize

economies of large scale which enable the utility to provide water to all users at a lower per unit cost.[17] However, this conclusion rests on very dubious reasoning, especially when the declining block rates encourage use during the peak or in a general situation of a shortage of capacity.

An Evaluation of Current Practices

The applications of rate principles by water rate analysts appear to reach an implicit compromise between "equity" and "efficiency." However, this compromise has resulted in practices which have led to economic inefficiencies and distributional impacts of a questionable character which are not explicitly recognized by utility practitioners.

The Concepts of Equity and Efficiency

The definition of equity employed in rate-making theory coincides roughly with the economic definition of efficiency. Given the underlying assumptions, economic efficiency is reached when price is equal to marginal cost. Rate differentials reflecting cost differentials at the margin satisfy this criterion.

Although the economic concept of efficiency roughly matches the utility concept of equity, the same relationship does not hold for utility "equity" and economic "equity." The economic definition of equity relates to the

---

[17] Howson, "Review of Ratemaking Theories," p. 849.

distributional problem. To be equitable in an economic framework, the rate structure does not have to reflect the total costs imposed by each user. Indeed, the rate structure may deliberately violate the marginal cost criterion, or any other criteria, if equity demands such violation. Therefore, to satisfy the criteria for economic equity, the rate structure does not have to be such that prices are equal to costs, and total revenue does not have to equal total costs. Revenue may be greater or less than costs, and the price may be greater or less than marginal cost.

Adherence to the utility concept of efficiency can lead to economic inefficiency. The tendency is for the utility solution to be based upon average costs and not marginal costs.[18] The case of a monopoly firm operating with constant factor prices, but realizing economies and diseconomies of size, where a divergence between average marginal costs occurs, is shown in Figures XXI_A and XXI_B below. In the case of decreasing costs, marginal cost is below average cost. If prices are based

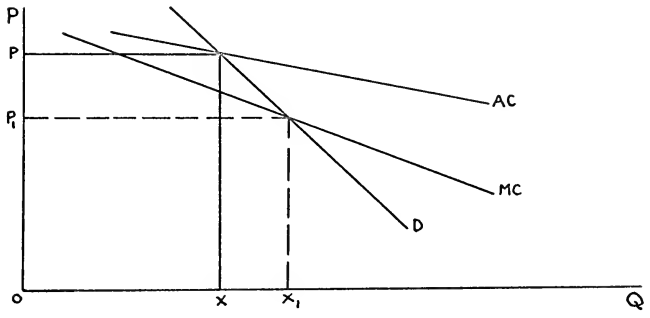
---

[18]The problem of average cost pricing as applied to water is discussed in the following: Hirshleifer, et al., Water Supply, p. 94. The writers state: "...where joint cost problems exist, an average-cost solution must then be objectionable in one of the following ways. First, if differing prices are charged because of cost allocations not proportional to use, the principle of equi-marginal value is violated. If the result is avoided by using proportional cost allocations and, consequently, a common price, the marginal cost principle will in general be violated." For other means of measuring and allocating costs see: J.C. Bonbright, "Fully Distributed Costs in Utility Rate Making," American Economic Review, LI (May, 1961), pp. 305-12; John R. Meyer and Gerald Kraft, "The Evaluation of Statistical Costing Techniques as Applied in the Transportation Industry," American Economic Review, LI (May, 1961), pp. 313-34; Milliman, "New Price Policies for Municipal Water Service," p. 129.

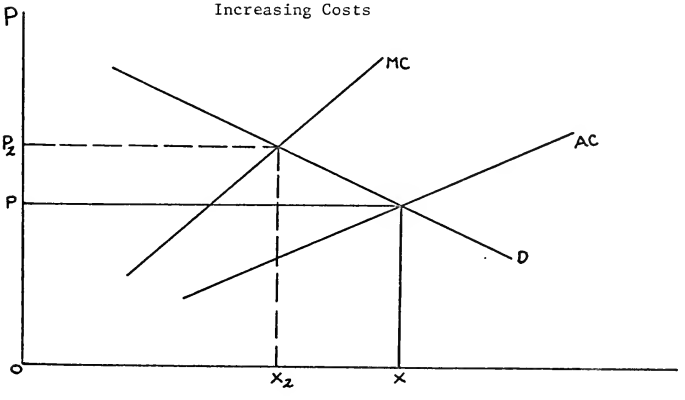
Figure XXI

Long-Run Costs of Monopoly Firm Producing
Water with Constant Factor Prices

A
Decreasing Costs



B
Increasing Costs



upon average cost, price is greater than marginal cost, and welfare is not maximized. The value of the marginal unit in the market is not equal to its value in production, and total welfare can be increased by reducing the price to P_1 (equal to marginal cost) and selling the output OX_1. Average cost pricing results in an output too small to achieve economic efficiency.

In the case of increasing costs (in Figure XXI_B above), setting price equal to average cost results in a price too low and an output too high to achieve the welfare maximum. Welfare can be increased by increasing the price to OP_2 and reducing output to OX_2. The losses arising from decreasing costs can only exist in the short-run. In the long-run, the firm may contract capacity and realize economies from a smaller plant, and convert the losses into a breakeven situation. In the case of constant costs, marginal cost and average cost pricing yield the same price and output since the two are equal.

Due to its adherence to average costs, the utility concept of efficiency can be roughly equated with the economic concept of equity, since the use of average cost pricing entails redistributions of surplus under assumptions of increasing and decreasing costs. In the case of decreasing costs, surplus is redistributed from consumers to producers, and from producers to consumers in the case of increasing costs. The redistribution results from the use of average cost pricing which in itself is inefficient.

An Evaluation of the Cost Allocation Technique

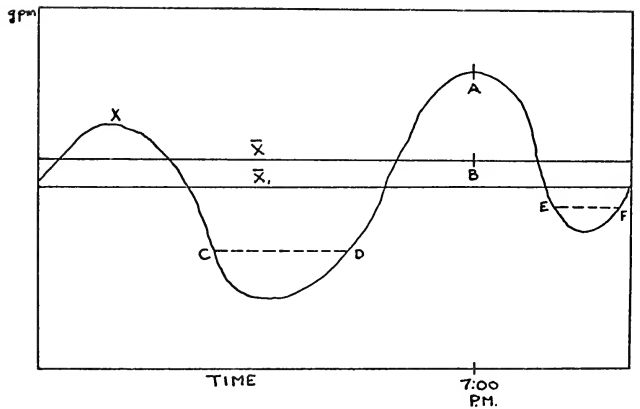
One of the basic problems the utility faces in the allocation of capacity costs (both production and distribution) is that of joint costs.

The one basic type of jointness which applies to the capacity costs is that of jointness between time periods. The attempt to allocate costs between base water costs and extra-capacity costs, and between customer classes, on an "equitable" basis comes about because of the time jointness between the peak and off-peak periods. It can be demonstrated that, if the time jointness could be eliminated, the cost allocation problem could be eliminated. However, the methodology employed by Patterson, as well as the water utility theory in general, would still indicate that an allocation problem existed. The utility theory would still have a problem because the use of customer classes and capacity classes entails the treatment of common or composite costs as joint costs. Since costs are allocated on the basis of the ratio of peak to average demand, a cost allocation would still be carried out, even though the problem associated with peak demand no longer exists.

Water utility theory generally, as well as Patterson, advocates the distribution of the system capacity costs between the base water costs and extra-capacity cost components on the basis of the ratio of the system's peak demand to its average demand. If the ratio is 4-1, base water costs bear 25 per cent of the capacity costs, and extra-capacity costs bear 75 per cent. When user classes are distinguished, the allocation of the total costs between classes is based upon the ratio between peak demand and average demand of each class. The use of demand ratios makes the cost allocation a pure demand allocation which can be affected by the use characteristics of the various customers. The problem is illustrated in Figure XXII where the solid line (X) is the actual

Figure XXII

An Hourly Load Chart



water demand for each hour of the day and line ($\bar{X}$) is the average demand. The ratio between points A and B determines the total capacity costs components of the base water costs and the extra-capacity costs. If two users, whose loads are shown by the dotted lines CD and EF, are added to the system (i.e., each will demand less than the system average), then the overall average will decline to line $\bar{X}_1$. As the average demand decreases, the ratio between the peak and average demand increases. Consequently, as the ratio increases, part of the base water costs are reallocated to extra-capacity costs. This reallocation is made even though there has been no change in peak water use, and no additional capacity has been required. The two additional customers utilize only excess capacity. The ratio between minimum and maximum use has decreased which would

indicate a reduction in the peak load problem, but peak responsibility has been increased. The peak users are now paying costs which they do not create.

Using averages to achieve "equity" brings this solution into conflict with the utility concept of equity when the argument is pushed to its extreme. Allocation of the capacity costs between various user classes--on the basis of their ratio of average and peak demands--uses peaks which are non-coincidental.[19] Non-coincidental peaks are used so that all users make a contribution to the extra-capacity costs. The position that the use of this ratio enables the rate structure to impose part of the capacity costs on all users, regardless of when water is used, is substantiated by Patterson:

> Allocation of extra-capacity costs among the customer classes is based upon non-coincidental extra-capacity requirements. This method of allocation most equitably distributes costs among customer classes whose individual maximum demands would not necessarily occur simultaneously, or at the time of the system peak.[20]

The allocation tends to penalize those users who do not take water during the system peak, or those users whose peak demand occurs at some time other than the time of the system peak. In Figure XXII, the system peak occurs at 7:00 P.M. If a user begins to take water for the three-hour period of 1:00 to 4:00 P.M., with a peak at 2:00, he shares in the extra-capacity costs. Since this user demands water for only three hours

---

[19] Patterson, "Practical Water Rate Determination," p. 909.

[20] Ibid., p. 909.

a day, his average hourly demand is very low, making his ratio between peak and average demands high. With a poor load factor indicated, the user pays a relatively larger share of the system's capacity costs than the user who has a lower ratio, but uses water during the system peak. Since the user in question does not have a demand for water during the system peak, he is not putting any burden on the utility's capacity. He is utilizing only excess capacity.

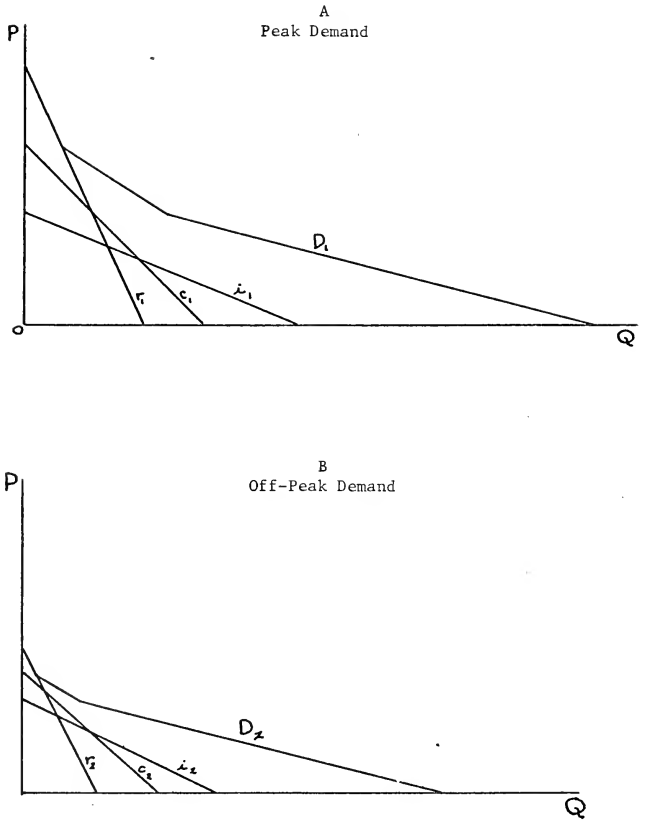
The ratio method of cost allocation tends to favor those users who consume water more hours than other users (who consume water fewer hours), since the former have a higher average hourly use than the latter. With a higher average and a given peak, the ratio between peak demand and average demand is lower. The allocation of costs based upon this ratio, therefore, reflects the difference in the demand characteristics of the users and not the costs imposed by the users. When rate differentials fail to reflect cost differentials, the cost allocation does not conform to the utility criterion for "equity."

In brief, the use of average and maximum demands tends to lead to a cost allocation that does not conform to the equity criterion the models themselves set up. The effect appears to be an allocation of costs based upon demand characteristics. As the demands change, the ratio between maximum and average demands might change, which, in turn, alters the division of total costs between the base water and extra-capacity cost components. The overall effect may very well be to build a series of distributional effects into the rate structure which lead to departures from economic efficiency and utility equity.

The use of averages and ratios to determine the cost allocation between base water and extra-capacity cost components, and between customer classes, as stated earlier, stems from the joint cost conditions between the peak and off-peak periods. The utility solution, however, seems to address itself to the wrong problem. The peak demand and the off-peak demand each represents a horizontal summation (not a vertical summation) of each of the demands in that period. In Figure $XXIII_A$ below, D_1 is the peak demand which is the horizontal sum of r_1, c_1, and i_1--the individual demands of the period. If time jointness did not exist, the utility could adjust its capacity to each of the demands, charge a single price to all users in each of the periods, and each user would contribute to the total costs in proportion to the amount of water taken in each of the periods. However, when time jointness exists, D_1 and D_2 (in Figures $XXIII_A$ and $XXIII_B$) are summed vertically and the costs are distributed on the basis of the relative strengths of the demands. Using a cost allocation entailing the use of averages and ratios, even when time jointness does not exist, requires allocating total costs between the base water and extra-capacity components, and between customer classes. The allocation is still required because the elimination of the time jointness (the ability to adjust capacity) does not eliminate the peak and off-peak variations in water use. Consequently, the methodology used by water engineers still has to grapple with its own allocation problem, even when the basic problem of jointness does not exist. The use of peak and off-peak prices in a manner prescribed by the economic models provides the solution based on an analysis of the problem as it exists--

Figure XXIII

Determination of the Peak
and Off-Peak Demands



time jointness between the peak and off-peak periods.

The general approach to cost allocation appears to be one of treating all cost allocations as joint cost problems. It is this treatment which appears to lead rate analysts to the conclusion that separate schedules should be developed for each customer class with costs allocated on the basis of the ratio of maximum demand to average demand. However, there is no justification for charging different users in the same time period different rates for the same quantity of service, since the costs are common costs and not joint costs. The decision to provide more capacity to serve residential customers during the peak does not enhance the utility's ability to serve commercial and industrial users. By the same token, the decision to serve more residential users with a given plant size reduces the utility's ability to serve the other classes. Thus, the costs, being common and not joint,[21] do not justify any price differentials between customers. Any difference is discrimination based on criteria other than economic efficiency criteria.

The Patterson model also ignores the impact of rate changes upon system design. His model was used to demonstrate how rates should be increased to make revenues equal to costs. The residential rate was increased by 27 per cent and the industrial rate by 50 per cent. However, Patterson ignores the impact that these higher rates might have on water use. It seems that Patterson implicitly assumes the demand to be

---

[21]For the distinction between joint and common costs see: Donald H. Wallace, "Joint Supply and Overhead Costs and Railway Rate Policy," Quarterly Journal of Economics, XLVIII (August, 1934), pp. 583-619.

completely inelastic, which would enable him to pass over a major part of the problem. However, the price charged is a determinant of the quantity of water used. The quantity used determines the size of the plant which, in turn, determines the costs that determine the rate the utility must charge to keep total revenue equal to total costs. It is difficult to accept the position that changing rates an average of 30 per cent will have no effect upon the quantity of water used, when it has been demonstrated (and cited in the following chapter) that water use is relatively sensitive to price changes.[22]

Summary of Cost Allocation Evaluation

There appears to be very little use of rate differentials between the peak and off-peak periods in water utility rate practices. There is recognition of the peak load problem through the allocation of extra-capacity costs. But the methodology fails to come to grips with the problem, as it tends to treat common costs as joint costs. However, the extra-capacity costs are built into the entire rate structure. Consequently, users having stable load factors (using the same amount of water as users with unstable load factors) pay the same total costs for the water used. The users with unstable loads are subsidized by users with stable loads, and the allocation of resources is distorted.[23] Individuals

---

[22]Gordon P. Fisher, "New Look at Resources Policy," Journal of the American Water Works Association, LVII (March, 1965), p. 259; J. Hirshleifer and J.W. Milliman, "Urban Water Supply: A Second Look," American Economic Review, LVII (May, 1967), p. 174.

[23]See: Bonbright, "Fully Distributed Costs in Utility Rate Making," p. 309; Milliman, "New Price Policies for Municipal Water Service," p.

not using water during the system peak pay part of the costs of serving the peak. Price is greater or less than marginal cost and total welfare could be increased by rearranging the prices so that the extra-capacity costs are placed on the users whose consumption of water coincides with the system peak.[24]

The allocation of resources appears to be distorted by the present utility practice of allocating costs on the basis of thoroughgoing averages. Average costs are allocated on the basis of average demands. Water rates are based upon average hourly and monthly use in spite of the fact that costs are determined by peak hourly flows. The apparent failure to come to grips with the true joint problem seems to have led to inefficiencies in the cost allocation serving as the basis for water rate structures.

An Evaluation of Water Charges

The use of various types of charges by a utility are designed to recover the costs incurred by the utility in providing water service. There is some question about the ability of these charges to satisfy the criteria for economic efficiency. When efficiency is not achieved, there

---

129; J.R. Nelson, "Practical Applications of Marginal Cost Pricing in the Public Utility Field," American Economic Review, LIII (May, 1963), p. 476.

[24]The bias towards not charging the peak user the entire burden is summarized by M.P. Hatcher who advocates the use of the demand charge. Hatcher feels that the use of the hourly peak puts too much burden on the residential user. M.P. Hatcher, "Basis for Rates," Journal of the American Water Works Association, LVII (March, 1965), pp. 273-78. This notion fails to incorporate the fact that the residential user is the greatest single source of the peak load problem.

are further distributional effects related to the departure from efficiency.

The use of fixed monthly charges may lead to serious efficiency problems. When users pay a fixed amount each month, the price might not be equated with the marginal cost of production. Also, there is no guarantee that water is being received by those individuals for whom it has the greatest marginal value (benefit). Since the cost of water is fixed to the user, the effect of the charge is the same as a lump-sum tax in that it does not alter the user's behavior, and might possibly lead to the inefficient and wasteful use of resources. Plant might be over expanded because there is no output rationing.[25]

The use of a flat charge per gallon has some of the same effects as the flat monthly charge, but it also improves upon some of the weaknesses of the latter. Since this rate is on a per-unit basis, the exchange value at the margin can possibly be equated for all users, which is not the case when a flat monthly charge is levied. This equating enhances the attainment of economic efficiency. However, the rate necessarily does not reflect the cost differences among the various users, and there is no guarantee that the exchange value at the margin is equal to

---

[25]However, if the costs of metering are great, the fixed charge can be the most efficient alternative. Metering is required only as the means of achieving efficiency prices since the cost a user imposes upon the system is directly related to the amoung of water use. But, if the costs of metering exceed the increase in efficiency which results from metering water use, the more efficient solution is not to use the meters since the total costs of supplying any given quantity of a given quality are increased.

the transformation cost at the margin. Consequently, the failure to achieve this equality leads to an inefficient use of resources. As inefficiencies develop, distributional effects arise. Since all users make a contribution to the capacity costs in proportion to their water use, users with good loads subsidize users with poor loads. Users with poor loads pay a price less than the marginal costs of their service while the users with a good load pay a price in excess of the marginal costs of service since there is no rate differential based upon the time and rate of water use.

The use of the minimum charge is designed to recover all customer costs, part of the base water costs, and part of the extra-capacity costs. In practice, the minimum charge results in a fixed charge for most residential users and contains the same distributional implications associated with the fixed monthly charge. Recent studies reveal that approximately 50 per cent of all water bills are for less than 3,000 gallons per month, and 73 per cent are for less than 5,000 gallons per month.[26] Since residential users constitute the greatest percentage of total users, present utility practice permits the greatest percentage of users to pay a flat monthly charge. The use of this type of minimum, when combined with meter costs, departs from economic efficiency and entails inter-class subsidies. The residential users comprise the class with the poorest load. Consequently, they probably receive a subsidy from those users with the better loads--most of the commercial and

---

[26]Albert P. Learned, "Determination of Municipal Water Rates," Journal of the American Water Works Association, XLIX (February, 1957), p. 169.

industrial users. Also, users in the immediate city tend to have more stable loads than users in the suburbs because of the lower incidence of lawn sprinkling. Since one rate is usually applied to all residential users, the city users pay part of the costs created by the suburban dwellers.[27]

The use of the service charge has been a means of overcoming the difficulty associated with the minimum charge. The users pay a flat monthly charge which includes all customer costs, and part of the capacity costs. However, there is no water allowance in the charge. All water use is metered and is charged through a variable use charge. This combination prevents any user from receiving water under a flat monthly charge. By using a variable charge, the water rate reflects the marginal cost of water which leads to greater economic efficiency. However, the service charge does not appear to be in broad use because users have come to expect a minimum water allowance in the charge. But this prejudice should not become a deterrent to greater adoption of rate schedules which eliminate subtle subsidies of which the various users are unaware and may not approve.

---

[27] Some communities attempt to correct the problem by charging a higher rate to outside users than to inside users. The outside rate differential is usually arbitrary and can range from 1.25 to 2 times the inside rate. The use of the city limits is arbitrary and can result in rate differentials which are not equal to the actual cost differentials. This solution does not correct the problem of the city dweller versus the suburban dweller when both live within the city limits. This general problem is discussed in a later section of the present chapter.

Of all the types of fixed charges used by various utilities, the charge which conforms closest to economic efficiency criteria is the tap-on charge. These costs are easily defined and can be assigned to specific users. These charges are based primarily upon the costs incurred in the installation of meters and the connection of the water lines.

Water rate analysts have been making progress in the direction of establishing rate structures which more accurately reflect the costs imposed by the various classes of users through the use of a minimum charge combined with a declining block-rate schedule. The use of the declining block rates has been justified on the basis that the more water that is used, the greater are the economies of scale the utility can realize. However, the argument has been pushed to the point where sales to large users at below cost prices in the last block have been justified on cost grounds. It is argued that the existence of the large users enables all users to realize lower prices; therefore, the large users deserve a subsidy from the small users for the attainment of efficiency.[28] It should also be noted that economic efficiency requires all users to pay the same price at the margin. Even in the case of decreasing costs, it might be argued that declining blocks do not lead to an efficient solution. The marginal costs will be determined by the level of total output. If there are two users, with one taking 5,000 gallons per month,

---

[28] Arthur Rynder, "Demand Rates and Metering Equipment in Milwaukee," Journal of the American Water Works Association, LII (October, 1960), p. 1240.

and the other 10,000 gallons per month, the former will pay a high per-unit price than the latter under a declining block-rate structure. However, the utility is producing 15,000 gallons per month, and the marginal costs are determined by this total output. Consequently, each of the users should be paying the same rate since they both contribute to the total demand (a horizontal sum) and are both marginal.

The service charge plus a declining block-rate schedule is a rate combination designed to recover both base water costs and extra-capacity costs in accordance with both the quantity of water used and the rate of water use.[29] However, due to the present expense of metering water to measure hourly use in order to arrive at peak contribution, the rate of water use is reflected as part of the service charge determined by the size of the user's meter. The users are, therefore, paying capacity costs in accordance with their ability to take water. Consequently, a user taking small amounts of water is paying a higher effective per-unit price than the user taking large amounts of water, if each has the same size meter. Unless this type of service charge is varied between user classes, such as the peak and off-peak user, there is a system of subsidies built into this type of charge.

---

[29]The rate of water use is collected through a demand commodity charge which is a charge that "...incorporates one charge based on the minimum rate of water use over a given period, usually an hour or a day, plus a charge for the quantity of water used," Keller, "Design of Water Rates," p. 294.

The Lack of Zone Pricing

The failure to use zone pricing leads to further departures from efficiency and greater distributional implications.[30] At an equal distance from the plant site, it is cheaper (on a per-unit basis) for the utility to serve densely populated areas than it is to serve sparsely populated areas primarily due to the number of connections per mile of water main and economies of larger operations. Consequently, people living in a densely populated area pay a price greater than their costs of service while users in more rural areas pay a price less than their costs of service.

There is one case where some distinction is made, particularly by the larger municipalities. The inside users pay a lower rate than the outside users. However, the usual distinction between the inside and outside users is made on the basis of the political boundries of the community. There appears to be very little effort to make distinctions in terms of economic or cost zones. High areas, where distribution costs are high, pay the same rate as low areas, where distribution costs may be nominal.

The use of arbitrary political boundries for rate-making can be justified on cost grounds if it is applied properly. The residents of the city can realize a return on their utility investment through either low rates or a reduction in other taxes in proportion to the contribution

---

[30] A model which develops this point has been presented by: Laurence C. Rosenberg, "Natural-Gas-Pipeline Rate Regulations: Marginal Cost Pricing and the Zone Allocation Problem," Journal of Political Economy, LXXV (April, 1967), pp. 159-68.

the utility makes to the city's general fund. In the case of the outside user, the utility is providing service to "non-owners" who impose additional costs upon the utility. The community is entitled to earn a fair return on this additional investment, and, since the outside users cannot be reached through other forms of taxation, higher rates to these "non-owners" is justified as the means of earning the return on the additional investment.[31]

The validity of this argument can be upheld only when these rate differentials are compared to the cost-of-service differentials. If the water plant is located on the city-county line, users in the county adjacent to the water plant might be paying a higher rate and a net subsidy to the city users across town who might be receiving water at rates below marginal costs. The failure to establish zones based upon economic criteria can lead to inefficiencies created by users receiving water at rates not in accordance with costs of service, and these departures from efficiency lead to distributional effects in the form of implicit subsidies.

Conclusions

The inability of utility literature to define rate structures and plant capacity stems from the bias towards average cost pricing, and the

---

[31]Samuel S. Baxter, "Principles of Rate Making for Publically [sic] Owned Utilities," Journal of the American Water Works Association, LII (October, 1960), p. 1237.

apparent failure to come to grips directly with the time jointness between the peak and off-peak periods. Another source of difficulty is the use of averages and ratios which results in the treatment of common costs (overhead) as joint costs.

The attempt to distribute costs between extra-capacity costs and base water costs, and between user classes, seems to bring out this joint cost treatment of the common costs. The effect is economic inefficiency since economic theory suggests that all users pay a single price in a common cost situation. The use of averages leads to a system of subsidies where the users with stable demands pay a price in excess of their marginal costs while the users with unstable demands pay a price less than their marginal costs of service. Users who take water only during the off-peak pay part of the costs of providing service to the peak water users. The failure to use zone pricing also leads to regional subsidies since the costs of serving all areas are unlikely to be the same.

The problems faced by a municipally owned utility also create difficulties not faced by privately owned utilities. An adequate summary of these problems is presented by G.P. Fisher, who states: (1) political and administrative decisions prevail over economic decisions; (2) it is easier to think in terms of expanding service than rationing through efficiency prices; (3) it seldom appears to public officials that price adjustments based upon rational economic analysis are possible; and (4) market processes are complicated by the non-market character of externalities.[32]

---

[32]Fisher, "New Look at Resources Policy," p. 259.

If these difficulties could be corrected, the problem of determining rate structures could become easier to resolve. R.J. Faust states:

> Too often town councils base their water rates on the schedules of neighboring towns, with no relation to the facts of local operations. It is no wonder that some water systems operate on insufficient funds and are unable to provide adequate service.[33]

At the same time, the failure of economists to recognize the needs and rationales of local government officials has been the source of much of the criticism the economist has had about the operation of municipally owned water utilities. There is a definite need for the application of economic analysis to the problems local governments are trying to solve through their water utilities. These problems include such factors as revenue sources, community growth and development, and distributional equity with respect to output and tax incidence.

---

[33]Raymond J. Faust, "The Needs of Water Utilities," Journal of the American Water Works Association, LI (June, 1959), p. 703.

CHAPTER V

APPLIED MODELS

Introduction

The purpose of this chapter is to explore rational alternative economic bases for determining water rates and plant capacity. The intent is to expand and supplement the theoretical models discussed in the preceding chapters so that they might have greater relevance at the level of application.

There is considerable criticism of water rate policies employed by water utilities. The critics maintain that if water rates were established on the basis of more sophisticated criteria, there would be a significant readjustment in the amount of water service presently being made available.[1] There is a tendency among engineers and long-range urban planners to project the "needs" for local service facilities without reference to service rate structures.[2] Implicit in such projections is the continued use of pricing systems which may not have a sound economic rationale. Obviously, conclusions as to future needs projected independently of future prices may be in serious error. It is possible that appropriate rate adjustments will ameliorate the problem. At least some

---

[1]Gordon P. Fisher, "New Look at Resources Policy," Journal of the American Water Works Association, LVII (March, 1965), p. 259.

[2]For an example of such planning see: Holly Cornell, "Comprehensive Water System Plan," Journal of the American Water Works Association, LX (February, 1968), pp. 125-28.

of the shortages would tend to disappear and arguments for new and larger facilities would be viewed within a more economic framework.

A second criticism of municipally owned utilities strikes at the social objectives they attempt to carry out through their rate structures. As stated by Irving K. Fox and Orric C. Herfindahl:

> It is extremely doubtful that the income redistributive consequences of existing subsidy provisions achieve any clear social objective. If the direct beneficiaries were required to pay for the services they receive, political support would more accurately reflect their social value.[3]

The basic approach to economic efficiency suggested in the economic literature is based upon marginal cost pricing. The recent analysis developed by Oliver E. Williamson sets forth a model that maximizes total surplus in a partial equilibrium model. However, even casual observation reveals that maximization in the Williamson sense is not the only, or even the dominant, objective at the applied level. It is necessary, therefore, to extend this model in order to assess the implications of profit maximization, price discrimination, and other objectives which are exceedingly important at the applied level. Although the economist cannot specify a policy objective, or determine the weight that should be given to competing objectives, he can assess the implications of alternativ objectives and suggest means for the attainment of given objectives.

---

[3]Irving K. Fox and Orric C. Herfindahl, "Attainment of Efficiency in Satisfying Demands for Water Resources," American Economic Review, LIV (May, 1964), p. 205.

Extension of the Model

The Williamson model equates "long-run" demand with long-run marginal cost to determine plant size and employs efficiency pricing to ration output in the short-run. During the off-peak, prices are equated to short-run marginal money costs and remain at this level so long as there is unused capacity. If the off-peak quantity demanded, at a price equal to short-run marginal cost, exceeds capacity output, price is increased until the quantity demanded becomes equal to capacity output. The conditions which satisfy a welfare maximum in the Williamson model are the same conditions which will emerge under perfect competition.[4] The marginal conditions which maximize welfare are invariant to the market structure, but are achieved under competitive conditions without the need for direct price and output determination at the level of public policy.

Assuming the peak and off-peak demands each last for twelve hours, the equilibrium solution is shown in Figure XXIV where the weighted sum of the peak and off-peak prices is equal to the long-run marginal cost. Total revenue is equal to total cost, making long-run profits equal to zero. The short-run marginal profit, P_1CDE, earned during the peak, just offsets the short-run loss, EDFb, incurred during the off-peak. These

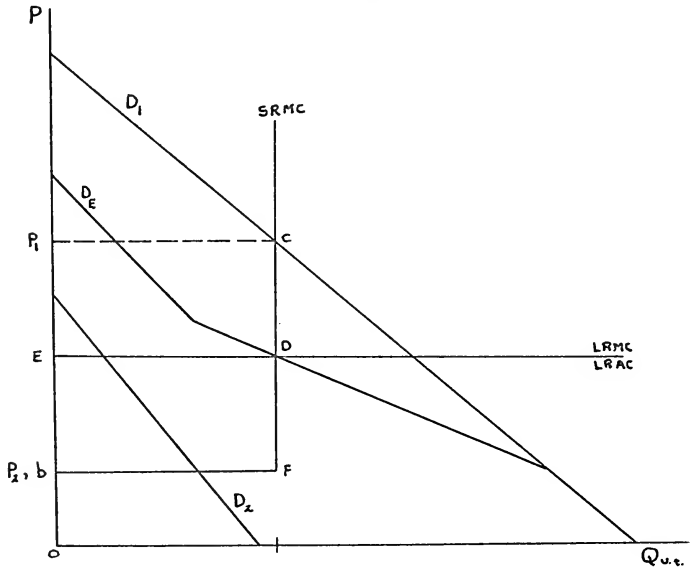
---

[4]The use of competitive assumptions to generate a solution similar to the Williamson solution when a firm faces a peak load problem is developed by William S. Vickery. Vickery's analysis is applied to a resort hotel and is, in essence, based upon a weighted summation of the peak and off-peak demands. William S. Vickery, Microstatics (New York: Harcourt, Brace and World, Inc., 1964), pp. 225-44.

conclusions are based on the assumption of a single price during the peak period and a single price during the off-peak period. The peak price, P_1, is charged for all units taken during the peak period, and P_2 is the price charged for all units taken during the off-peak period.

Figure XXIV

The Williamson Model

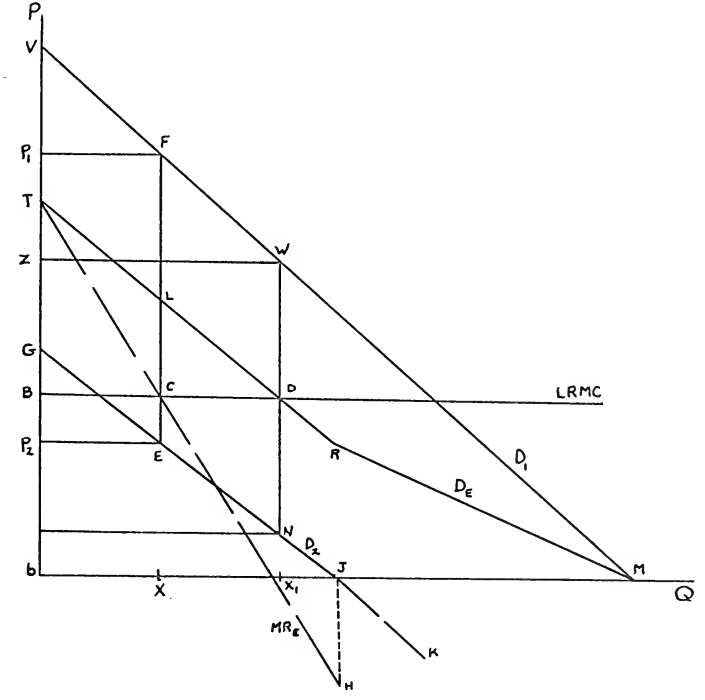


Retaining the assumptions of joint costs, independent demands, single prices, and constant costs, the basic model can be modified to depict a firm concerned with profit maximization rather than welfare

maximization in the Williamson sense. This objective might apply in the case of an unregulated private water utility, or in the case of a municipally owned water utility attempting to maximize net revenue. The profit-maximizing solution is shown in Figure XXV. The line TCHJK is the marginal function to the "long-run" average demand (TLDRM). Profits are maximized at output OX where long-run marginal revenue equals long-run

Figure XXV

Non-Discriminating Monopoly



marginal cost. The profit-maximizing plant size is defined by point C which, of course, is smaller than the welfare maximizing capacity defined by point D.[5] The peak price is bP_1, the off-peak price is bP_2, and total profit is equal to P_1BCF (peak profit) minus $BCEP_2$ (off-peak loss).

The monopoly arrangement, since it results in a smaller size plant than the welfare maximizing plant, leads to a net loss of welfare to users. When welfare is maximized, total surplus is equal to the quantity $VWX_1b + GNX_1b - 2(BDX_1b)$. However, if profits are maximized, total surplus is equal to VFXb + GEXb - 2(BCXb). Total welfare, as measured by total surplus, is reduced by a net amount equal to the triangle CLD when the firm pursues a profit-maximizing policy.

Presumably, regulation of privately owned utilities prevents water prices from reaching these monopoly levels. However, municipally owned utilities are not constrained in this way. On the assumption of a single price in each period, the model in Figure XXV is the one a community might tend to follow if the intent is to provide maximum general fund revenues. However, plant size C is the smallest that can be conceived on the basis of any reasonable assumption concerning the objectives of a utility.

---

[5]These comparisons assume that the firm charges a single price to all users in any given period. Although the price may differ between the peak and off-peak periods, all users pay the same peak and off-peak price. In other words, there is no rate distinction made between user classifications, only between time periods.

All the evidence seems to indicate that municipal water utilities operate with excess capacity during the off-peak period and do not use a single, or non-discriminatory, pricing system. It becomes obvious that the policy described in Figure XXV is not the one being pursued by either privately or publically owned water utilities. In the case of municipally owned water utilities, large monopoly profits can be "absorbed" in empire building which might lead to a plant size greater than that which maximizes profits. Excessive profits may also be employed to provide excessive staff and emoluments, or to provide political patronage in one form or another. These possibilities are enhanced if the utility need only make a given or "expected" contribution to the community's general fund revenue.[6] Thus, it is possible to have a monopoly price coupled with a plant size that is greater than that defined by point C or even that defined by point D. The effect on efficiency is ultimately measured by the deviation of the actual plant size from that size defined by point D. This effect is described in a subsequent section.

The explanation for policies not conforming to the policy described in Figure XXV must be found in the rate structures that are actually used. Utilities do not employ a single price nor do they employ peak and off-peak prices. Block prices applicable to all periods are the usual practice among the larger utilities. The use of block pricing leads to a solution which is different from the single price profit-maximizing

---

[6]The same possibility exists in the case of privately owned utilities if the regulatory authorities do not carefully scrutinize costs. See: Harvey Averch and Leland L. Johnson, "Behavior of the Firm under Regulatory Constraint," American Economic Review, LII (December, 1962), p. 1052.

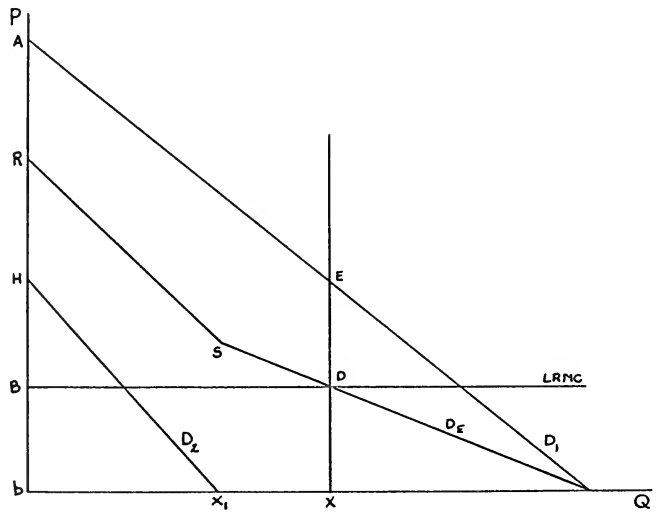
monopoly model summarized in Figure XXV.[7]

The Economics of Block Pricing

The effect on rates and plant size when a firm employs first degree price discrimination combined with the objective of profit maximization is shown in Figure XXVI. Assuming zero, or negligible, income effects from perfect first degree discrimination, the demand curve of the firm

Figure XXVI

First Degree Price Discrimination



[7]James M. Buchanan, "Peak Loads and Efficient Pricing:Comment," Quarterly Journal of Economics, LXXX (August, 1966), pp. 463-71.

becomes its marginal revenue curve.[8] Plant size is determined where the effective demand (marginal revenue) function D_E intersects long-run marginal cost. The firm charges all possible prices along the segment AE of the peak demand (D_1) and along the entire segment of the off-peak demand (D_2). Total revenue is equal to AEXb + HX_1b and total costs are equal to 2(BDXb). The plant size is defined by point D--the plant size generated under conditions of perfect competition! Total welfare, as measured by surplus, is equal to the area RSDB--the same as that under conditions of perfect competition! However, all consumers' surplus is converted into profits as each user buys each unit in accordance with the marginal benefit derived from each unit.[9] Thus, the basic difference between the perfectly discriminating firm and the non-discriminating firm revolves around questions of equity rather than efficiency.

An applied alternative to both a single price monopoly and complete first degree discrimination is the use of block pricing, which represents a combination of first and third degree discrimination. As an individual uses larger amounts of water, additional quantities can be purchased at decreasing prices, thus achieving first degree discrimination. The rates

---

[8]The assumption of negligible income effects simplifies the analysis. Without the assumption, the solution becomes indeterminate. This point is well developed by Buchanan, "Peak Loads and Efficient Pricing: Comment," pp. 463-71.

[9]This interpretation equates an individual's demand curve with a marginal benefit curve if income effects are assumed away or considered negligible. For this type of construction see: Lawrence Fouraker, "A Note on the Administration's Recent Tax Proposal," National Tax Journal, XVI (December, 1963), pp. 426-28; Vickery, Microstatics, p. 228.

between user classes can also be varied to attain third degree discrimination.[10] In this case, there may be a different rate charged to residential users than is charged to commercial users. A method of determining profit-maximizing rates using block-rate discrimination is shown in Figure XXVII, which assumes constant costs and a single demand.[11] A solution using three blocks entails breaking the segment AB of the demand curve, D, into four equal parts. Each part represents a potential block. The first block is quantity OX_1 sold at the price X_1P_1. The second block is equal to quantity X_1X_2 to be sold at the price X_2P_2. The third block quantity is X_2X_3 sold at the price X_3P_3. No additional units beyond quantity OX_3 are sold since, at the quantity OX_3, marginal revenue is equal to marginal cost. The sale of additional units beyond OX_3 entails losses on these units as marginal revenue falls below marginal cost. A quantity greater than OX_3 is sold only if the number of blocks is increased. The model constructed in Figure XXVII maximizes profits on the assumption that three blocks are to be employed. No other three-block schedule can be derived which yields as much profit as $(FP_1JH + GP_2KJ + EP_3LK)$.

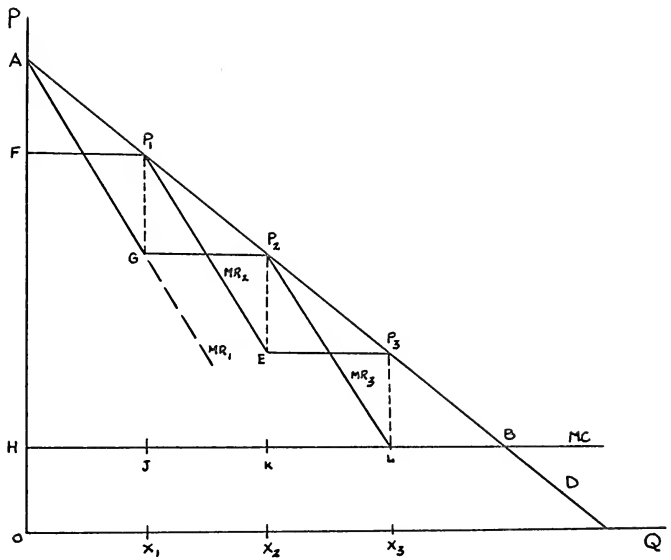
---

[10] It should be noted that third degree discrimination would be superfluous if it were possible to employ perfect first degree discrimination, but the latter requires an estimation of each individual's demand and a separate rate schedule for each individual (unless all demands are assumed to be identical). Block rates and user classifications are simply devices by which the market is segmented. First degree discrimination represents the ultimate in segmentation since each unit is sold to each customer at a different price.

[11] For a complete discussion of applied block-rate pricing see: Ralph K. Davidson, Price Discrimination in Selling Gas and Electricity (Baltimore: Johns Hopkins University Press, 1955), pp. 148-80.

Figure XXVII

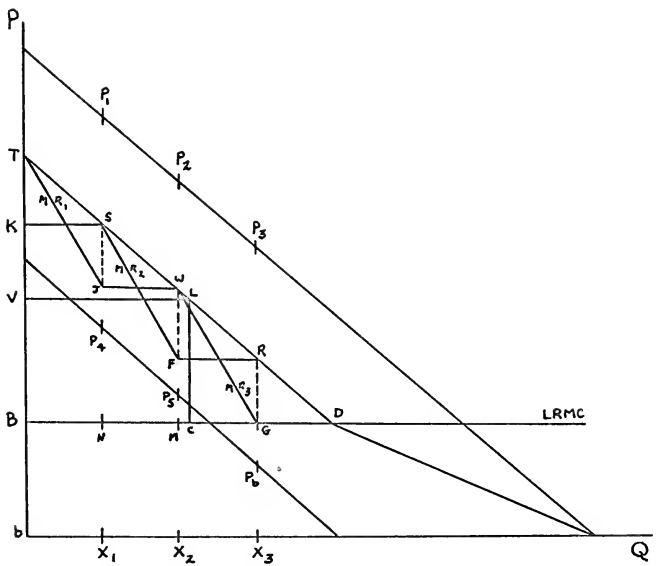
A Block-Rate Model



The usual solution for profit maximization under a block-rate structure rests upon the assumption of a single demand, that is, there is no peak and off-peak problem. The Williamson model is based upon the assumption of two demands which are independent and a single price for each of the two sub-periods. The two analyses can be brought together to demonstrate how block-rate pricing can be applied when the firm is faced with more than one demand in the same time horizon. The general model is shown in Figure XXVIII. Total output is determined by point G where

Figure XXVIII

Williamson's Model with Block Pricing



marginal revenue is equal to long-run marginal cost. The effective demand curve, D_E, becomes the relevant curve in the determination of the blocks. If three blocks are used, the segment, TD, is divided into four equal parts. The curve which becomes the relevant marginal curve to the "long-run" demand is that labeled MR_1, MR_2, and MR_3. At total output, bX_3, marginal revenue is equal to long-run marginal cost. The price of the last block during the peak is determined by that output which equates marginal revenue with marginal cost. The price of the last

block in the off-peak period is set where off-peak marginal revenue is equal to short-run marginal cost, b, which in this model is assumed to be zero. The capacity is sufficient to produce a total output of bX_3 which is sold during both the peak and the off-peak periods. Quantity bX_1 is sold in the first block at the price X_1P_1 during the peak period and price X_1P_4 during the off-peak period. The second block quantity is X_1X_2 and is sold at the price X_2P_2 during the peak period and X_2P_5 during the off-peak period. The third block quantity, X_2X_3, is sold at price X_3P_3 during the peak period and price X_3P_6 during the off-peak period.

The welfare implications of this alternative are different from the other solutions already discussed. By referring back to Figure XXVIII, the overall differences in the effects of each alternative in terms of efficiency can be seen. Under conditions of perfect competition and perfect first degree discrimination, plant size is that defined by point D, and total welfare, as measured by surplus, is equal to the area of the triangle TDB.[12] That is, if profits are maximized under a system of perfect price discrimination, and if income effects are negligible, the plant size and the distribution of the output among consumers will be optimal. The distribution of the surplus must rest upon distributive judgments and interpersonal comparisons. The single price monopoly profit-maximizing output defined by point C yields the lowest total surplus

[12]The consumers' surplus measured under the effective demand curve is the weighted sum of the surplus derived by the off-peak users during the off-peak period and the surplus derived by the peak users during the peak.

(TLCB) and the distribution is TLV as consumers' surplus and VLCB as producers' surplus. Total welfare using a three block-rate structure is equal to TRGB in which consumers' surplus is equal to TSK + SWJ + WRF, and producers' profits are equal to KSNB + JWMN + FRGM. The three blocks yield a plant which falls between points C and D, since TLCB is less than TRGB which is less than TDB. The extent to which a block-rate schedule yields a total surplus approaching the surplus yielded by the plant sizes defined by points C and D depends upon the number of blocks in the rate schedule--the greater the number of blocks, the closer total surplus will approach equality with TDB (maximum surplus), and the fewer the blocks, the closer total surplus will be to TLCB.

Models which Interpret Present Practices

The ability of municipally owned water utilities to establish profit levels and rate structures without the fear of regulatory sanctions can lead to a non-economic level of plant investment.[13] The knowledge that rates to small users with relatively inelastic demands (the captive market) can be increased without affecting water use to any great extent can lead to the expansion of production beyond the optimum. Sales at below marginal cost can be used to provide subsidies to special interest groups or to provide location incentives for industry. These subsidies can be

---

[13]This position does not imply that regulation will necessarily prevent such excesses. However, the types of sanctions and limits ordinarily associated with the regulation seem totally absent in the case of municipal utilities where there is a presumption that the interests of management and consumers are not in conflict.

offset by sales above marginal cost to certain other segments of the market.

When a utility is permitted to practice discrimination for purposes of recovering losses arising from decreasing costs, or from sales at below cost, there is no check to excessive discrimination. Indeed, there is no criterion for determining the proper amount of discrimination independently of the distributional question. However, "unjust" or unfair discrimination is a strong possibility in the case of municipally owned utilities where there is no direct regulation over the utilities' practices.

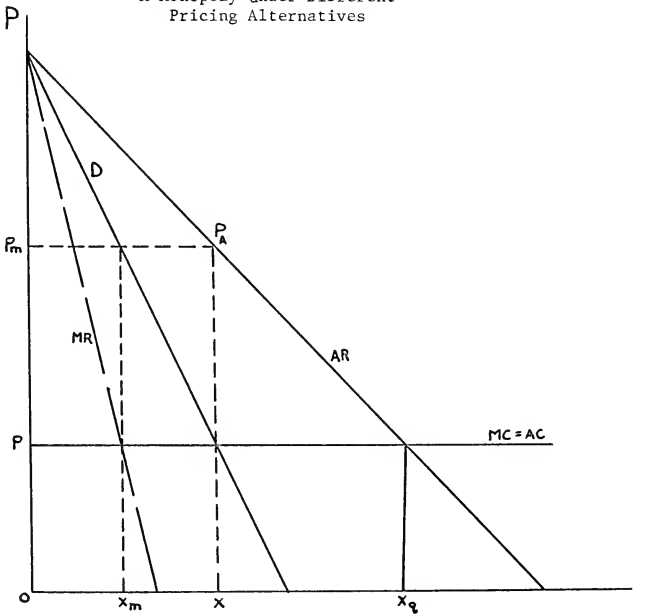
The models developed to this point can be useful in explaining, at least in part, some of the actual conditions and practices municipally owned water utilities tend to employ. It is clear that most water utilities do not maximize monopoly profits. Although some utilities make significant contributions to general revenue, others break even, and a number require general fund subsidy. However, price discrimination is practiced by virtually all utilities. The question that arises is: under what circumstances can a utility which discriminates in price develop only modest profits? Alternatively, what are the implications of discriminatory prices coupled with either a voluntary or imposed profit restraint?

In order to examine this problem in simple terms, assume a monopolist operates under conditions of constant costs and has no peak load

problem.[14] In Figure XXIX below, X and P are the welfare maximizing price-output relations under the assumption of a single price. The introduction of perfect price discrimination shifts marginal revenue

Figure XXIX

A Monopoly under Different Pricing Alternatives



[14] The analysis of pages 140-4 relies on a general treatment by Milton Z. Kafoglis, "Output of the Firm under an Earnings Restraint," unpublished manuscript (University of Tennessee, 1968).

to (D) and creates a new average revenue curve (AR). The discriminatory profit-maximizing monopolist now produces the optimal output OX at the average price P_A. In many instances, this solution, which is associated with optimal output, exhorbitant profits, and highly discriminatory prices, may be superior to the simple monopoly solution with restricted output and smaller profits since total welfare, as measured by surplus, is greatest at the optimal output. The discriminating monopolist sells to some customers at a higher price than the simple monopolist, to others at a lower price. The welfare choice, insofar as output is concerned, favors the discriminating monopolist; insofar as prices and profits are concerned, no judgment can be reached except on distributional grounds. Observation of utility practices suggests that discrimination has been the policy choice. However, the profits seem to have been dissipated except for the limit imposed by regulatory agencies in the case of privately owned utilities, or the restraint imposed by elected officials in the case of publicly owned utilities.[15]

The discriminating utility confronted with exhorbitant profits may (a) reduce the degree of discrimination, (b) inflate costs, and/or (c) increase output by selling output beyond OX (in Figure XXIX) at a price

---

[15]The restrictions imposed upon publicly owned water utilities are twofold. The general funds derived from the water utility are not usually an open-ended amount. The community's financial officers usually expect some set amount of revenue from the utility to be applied to the general fund. Also, when rates become such that the profits become excessive beyond this amount, political pressure mounts for a rate reduction.

below marginal cost. Any of these alternatives will use up the excess profits.

If discrimination is reduced, AR will shift to the left (Figure XXIX) as profits are eliminated. When profits are completely dissipated, AR will coincide with D, and the Pareto optimal output (X) will be obtained at the Pareto optimal price (P). This alternative requires a voluntary transfer of surplus from the producer (or the general fund) to the water consumer. Experience suggests that neither the management nor local officials are likely to adopt this alternative.

If discrimination is maintained, excess or unwanted profits may be used up in cost increasing activities including managerial emoluments, "services" to the community, and other fringe items. The management might develop a preference for expense which may be exceedingly difficult to control depending upon the amount of autonomy given to the utility officials. This "expense preference" model eliminates unseemly profits without restricting output in the traditional monopoly sense and without reducing prices. The income transfer is a transfer from consumers to management. Needless to say, political patronage and political machines can be developed on the assumption of expense preference.

The third possibility is to increase the degree of discrimination by reducing marginal rates to a level below marginal cost. Thus, in Figure XXIX, the utility can expand output to OX_q and eliminate profits completely by selling the segment $X\text{-}X_q$ at a price below marginal cost. The plant is larger than optimum, and welfare, as measured by surplus, is reduced. There is extreme over investment, and some customers are being charged what the traffic will bear while others are receiving bargain

rates. This alternative reflects a managerial preference for a larger empire and would seem to be a more efficient device for political purposes than the expense preference model. This alternative is also consistent with the "community development" objective through the subsidy of large industrial users at a price below marginal costs. It is consistent with the political practices of supplying some "important" commercial and industrial users with favorable rates, as well as providing the rationale behind low promotional rates and the alleged "economies" derived from serving large users. This model also has relevance for municipally owned utilities since the objectives of these utilities are not to acquire as much revenue as possible. Most community financial officers have a relatively fixed sum of money they expect to receive from the utility for the general fund revenues. When this sum of money is acquired, any additional utility net revenue is considered "undesirable." To avoid these excess revenues, the utility managers can carry out expansion of output at rates below cost to use up the revenue, such as the expansion into markets where the costs are greater than revenues.[16]

A Further Modification

An alternative model, which represents a variation of the quantity maximization or empire building model, is shown in Figure XXX below. This model considers the general approach of utility managers, which, from all

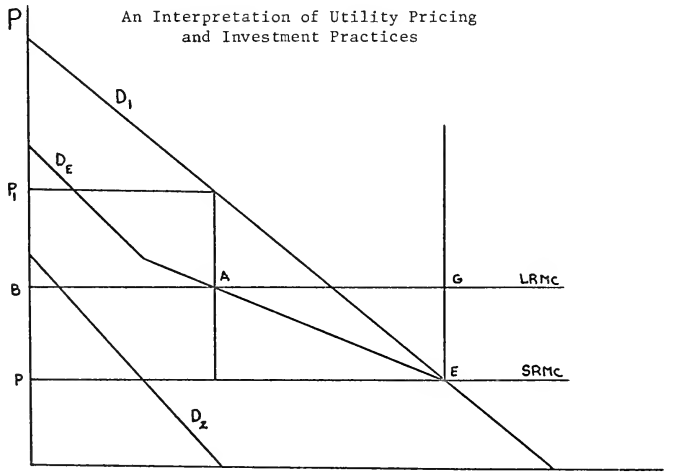
---

[16]Averch and Johnson, "Behavior of the Firm under Regulatory Constraint," p. 1058. Also see Oliver E. Williamson, The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm (Englewood Cliffs: Prentice-Hall, 1964); Milton Z. Kafoglis, "The Public Interest in Utility Rate Structures" (unpublished and undated manuscript).

indications, appears to be one of supplying the total quantity of water demanded at a price equal to short-run marginal cost (out-of-pocket). The only form of rationing appears to be the ability of water meters to handle a water flow at any given point in time.[17] The use of out-of-pocket cost pricing leads to a vast over expansion of water facilities. In Figure XXX, D_1 is the peak demand, D_2 the off-peak demand, and the utility operates under an assumption of constant costs. The utility charges the

Figure XXX

An Interpretation of Utility Pricing and Investment Practices



[17] This alternative appears to be indicative of utility practices because of the prevalent use of an open-ended block at the end of the rate structure rather than a closed block. Consequently, the utility must be prepared to supply any amount at the given price in the last block.

same marginal price (OP) to both the peak and off-peak users. The marginal price is equated with short-run marginal cost. Plant size is that defined by point E where the peak demand intersects the short-run marginal cost function. Using a single marginal cost price of OP, the firm takes losses equal to 2(BGEP). It is obvious that the plant is over expanded since the economic solution requires the plant capacity defined by point A to be constructed (where D_E intersects long-run marginal cost) with a peak price of OP_1 and an off-peak price of OP. The total loss in welfare from setting both prices equal to short-run marginal cost is equal to the area of the triangle AGE when compared to the optimal capacity defined by point A. The marginal solution is one of the worst possible alternatives for pricing and investment decisions. Since the utility takes losses by setting prices equal to short-run marginal costs, the use of discrimination on the infra-marginal units becomes the means by which the losses are recovered.

The ability to practice discrimination along with the fact that water utilities seem pressed continuously suggests that the alternative described in Figure XXX--quantity maximization and short-run marginal cost pricing--may be a major source of aggrevation in the municipal water supply picture.

Administrative Problems

The peak and off-peak models presented to this point have assumed that demands are independent of each other. It is this assumption that permits a differential between peak and off-peak rates to be used. The assumption is realistic when dealing with the problem of seasonal variations

in water use since users are unable to shift their water use from the peak summer months to the off-peak winter months. There is the possibility that storage facilities can be constructed and individual water users can store water during winter months for use during summer months when water rates are high. However, rate differentials would have to be extremely high to make this alternative profitable. If water could be stored in sufficient quantities by any given user, there remains the problem of keeping the water from becoming impure and stagnant. These latter factors prevent storage from becoming a significant alternative which merits further consideration. Thus, lawn sprinkling carried out during summer months cannot be shifted to winter months. Swimming pools used during the summer cannot be filled during the winter. Seasonal demands are, therefore, independent of each other and price differentials can be used between the peak and off-peak periods. However, this is not the case with respect to the hourly peak.

The demand for water on an hourly basis is much more sensitive to price differentials than is the seasonal demand for water since the two demands are no longer independent of each other.[18] The demand for water

---

[18] In a recent study by Charles E. Howe and F.P. Linaweaver, Jr., some significant factors were found with respect to the elasticities of demand for water. For the domestic demand for water, price elasticity was estimated to be -0.23. Income elasticity, measured by property values, was estimated to be 0.35. The analysis of summer sprinkling demand showed price elasticity in dry western areas to be -0.7 whereas in humid eastern areas, price elasticity was -1.5. In the analysis of maximum day sprinkling, the study found that in arid western areas, the magnitude of water use did not respond significantly to price changes. The amount of water used during the peak day in humid eastern areas did respond to price changes, but the elasticities were considerably below those

during the hourly peak period becomes a partial function of the price charged for water during the off-peak period. Therefore, the water utility which uses peak load pricing might be faced with the problem of continually changing rates. The previous peak period could become the off-peak period and the previous off-peak period could become the peak period if price differentials are maintained.[19] The problem of shifting peaks can be avoided through the use of a fixed charge. The same block schedule can be used for both periods, but a minimum charge with no water allowance can be used to develop the rate differential. Due to the expense of metering water to measure hourly use, a minimum fixed charge designed to collect the additional costs arising from the hourly peak demand can be based upon the size of the water meter the user selects--the larger the meter, the larger the service charge.

The fixed charge becomes one based upon the amount of water the user is capable of taking during the system peak. However, there are some equity problems which must be recognized in the use of this alternative. Those users who take no water during the system peak must pay part of the hourly peak capacity costs. The user, upon the selection of the water line

---

for the summer sprinkling demand. The income elasticities of demand were considerably lower than that of summer sprinkling demand. The income elasticities of the daily demand for lawn sprinkling are lower than that for the seasonal period. Charles E. Howe and F.P. Linaweaver, Jr., The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structures (Washington: Resources for the Future, Inc., 1967), pp. 27-29.

[19]This phenomenon has been a major problem in the telephone industry where continuous attempts to diminish the peak have resulted only in shifting the peak; leading to the chaos in rates which has become the subject of a significant and far-reaching investigation by the Federal Communications Commission.

needed for his purposes, is presented with the various size meters and their respective service charges. The user is then free to select the combination of meter size and service charge he desires which enables him to move to the preference level he desires. Due to present technological inefficiencies in water meters, a user cannot select a meter smaller than he needs in order to avoid a higher service charge, and then attempt to run his water taps wide open during the peak period. Water pressure will drop, and, at the same time, the meter will record more water than is actually flowing through it due to friction. Consequently, the user gets lower quality service and pays for more water than he actually uses.

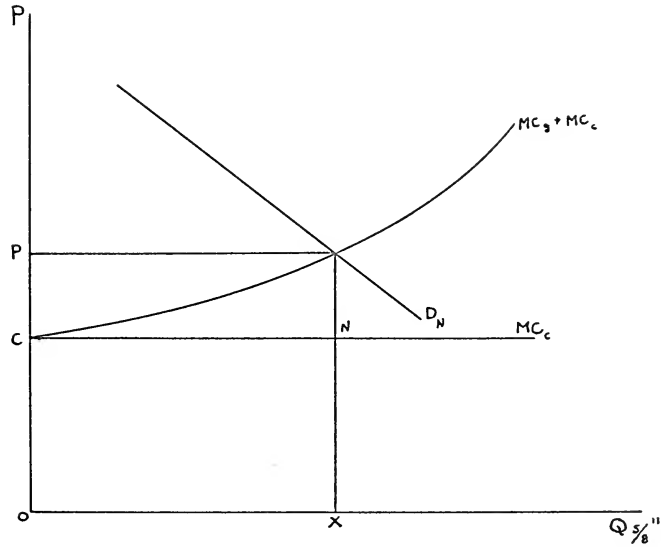
Because the capacity charge is fixed and not variable with water use, the user has no reason to shift his peak use to other hours. Therefore, the utility is able to recapture hourly peak costs and avoid the problem of a shifting peak. In using a fixed charge, all customers' costs and hourly peak costs are put into the fixed charge and the base water costs are allocated through a variable charge related to water use. The service charge is applicable to the various size meters and can be determined by the number of equivalent 5/8" meters (the basic meter) each meter represents.

A Fixed Charge Model

A model, using the ability to congest as the basis for a fixed charge to recover customer costs and hourly peak capacity costs, is shown in Figure XXXI below where D_N represents the "net" demand for water during the hourly peak. The demand is the net demand over and above the seasonal peak demand which can not be handled through a variable charge because of

Figure XXXI

Fixed Charge Based upon the Ability
to Congest the Water System



the interdependency of the hourly demands. The function MC_c is the marginal customer cost associated with serving an individual customer for each equivalent 5/8" meter. The function MC_g is the estimated marginal congestion costs created by an equivalent 5/8" meter during the hourly peak. These costs are due primarily to pressure losses arising from the heavy demand placed upon the system capacity. They can be converted into money costs if they are interpreted as the additional costs the utility must incur to offset pressure losses arising from different quantities

of equivalent 5/8" meters. Engineering studies should be able to make available the data necessary to obtain the estimated water use which will contribute to the hourly peak load problem when different quantities of equivalent 5/8" meters are supplied. The price-output relationship is OP price and OX quantity of equivalent 5/8" meters. The fixed charge is OP, with OC the part used to recover the marginal customer costs and CP the part designed to recover the capacity costs imposed by an equivalent 5/8" meter during the hourly peak. The charge OP might be imposed as a lump-sum tax payable at the time the meter is installed, or it might be converted into a monthly charge by dividing OP by the number of months in the estimated life of the investment. The application of the charge OP, as noted, is applied in terms of equivalent 5/8" meters. The household with a 5/8" meter pays a charge equal to OP. The user with a 1" meter, which might be the equivalent of three 5/8" meters, pays a charge equal to OC as the customer cost plus 3(CP) as the capacity charge.[20]

Some of the equity implications of this alternative should be noted. The congestion charge is based upon engineering data which reflects the amount of congestion associated with serving different quantities of equivalent 5/8" meters. Therefore, each customer with a meter of a given size is paying a charge based upon the ability of this user class to congest the water system. In this respect, the fixed charge is efficient

---

[20]The customer charge of OC can be altered in proportion to the change in costs associated with the installation and servicing different size meters. A 1" meter does not cost 3 times as much to install as a 5/8" meter, even though the 1" meter can contribute to system congestion 3 times as much as the congestion created by a 5/8" meter.

since all users pay the same fixed charge, given the meter size. However, those users not taking water during the system's hourly peak (or less water than their meter is capable of delivering) are paying part of the costs imposed by the users who take the maximum amount of water they can during the peak. Since residential users in the city create less of a peak load problem than residential users in the suburbs, in the absence of zone pricing, the city dwellers pay part of the costs created by the urban dwellers, and, hence, a "rate subsidy."

This pricing alternative provides a general solution to the problems of determining capacity and rate structures when demands are interdependent and external social costs exist. The social costs are converted into money costs as capacity is added to overcome them. The service charge takes the form of a lump-sum tax and reflects all the costs of providing water service when prices are equated with short-run marginal costs. Although the charge does nothing to eliminate the peak problem, it does prevent the problem of a shifting peak and recovers all the costs associated with the hourly peak use while allocating costs to those user classes who impose the capacity costs.[21] The seasonal peak and off-peak

---

[21]For excellent studies in the use of congestion pricing see the following articles: Clifton M. Grubs, "Theory of Spillover Cost Pricing," Highway Research Record, No. 47 (Washington: Highway Research Board, 1964), pp. 15-22; Herbert Mohring, "Relation Between Optimum Congestion Tolls and Present Highway User Charges," Highway Research Record, No. 47 (Washington: Highway Research Board, 1964), pp. 1-14; G.P. St. Clair, "Congestion Tolls-An Engineers Viewpoint," Highway Research Record, No. 47 (Washington: Highway Research Board, 1964), pp. 66-112; A.A. Walters, "The Theory and Measurement of Private and Social Costs of Highway Congestion," Econometrica, XXIX (October, 1961), pp. 676-99.

rates will be varied according to the months of the year. The costs of the seasonal peak demand will be recovered through the peak variable charge levied on the basis of a per-gallon charge. All the peak users will pay the same per-unit price during the peak period through the use of this type of variable charge.[22]

Also, it should be noted that this model results in a solution which is welfare maximizing in the Williamson sense. The basic difference is that the model in Figure XXXI assumes congestion is an increasing function of the quantity of equivalent 5/8" meters while Williamson assumes all costs are constant. The function $MC_g + MC_c$ is equal to Williamson's long-run marginal cost function.

Zone Pricing

One form of pricing which is lacking in many of the solutions to water utility problems is zone pricing. The general conclusions have been

---

[22]This solution eliminates the use of declining block-rate structures which are inefficient since their use entails different users paying a different per-unit price which is determined by the amount of water taken in any given period. A user who takes 5,000 gallons per month pays a higher marginal price than the user who takes 10,000 gallons per month under a block schedule. However, the marginal cost of production is determined by the total output of 15,000 gallons per month. Consequently, each of these users is a marginal user; hence, both should pay an equal marginal rate. For studies which make some recommendations about rate structures see: John Hopkinson, "On the Cost of Electric Supply," The Development of Scientific Rates for Electric Supply (Detroit: The Edison Illuminating Co., 1915), pp. 5-20; Arthur Wright, "Cost of Electricity Supply," The Development of Scientific Rates for Electric Supply (Detroit: The Edison Illuminating Co., 1915), pp. 31-52.

that,if there is excess capacity during the off-peak period, all capacity costs are to be paid by peak users. Since municipal water companies tend to operate with excess capacity during the off-peak, those conclusions suggest significant changes in rate practices. However, the allocation of all capacity costs among peak users may not be the most efficient solution. It can be argued that some users should pay part of the capacity costs even though they use water only during the system's off-peak period, particularly in the case of distribution costs.[23] It is important to examine this argument.

The analysis which is developed in this section is an adaptation of a model developed by Laurence C. Rosenberg.[24] Although Rosenberg applies his analysis to the allocation of costs associated with gas pipelines, it is readily adapted to the distribution costs of a water utility. Assume the utility is going to serve three zones in which Zone A is within a one mile radius of the water plant, Zone B is within a two mile radius of the plant, and Zone C is within a three mile radius. Marginal costs of serving each zone are shown in Figure XXXII and Table 1 below. Costs

---

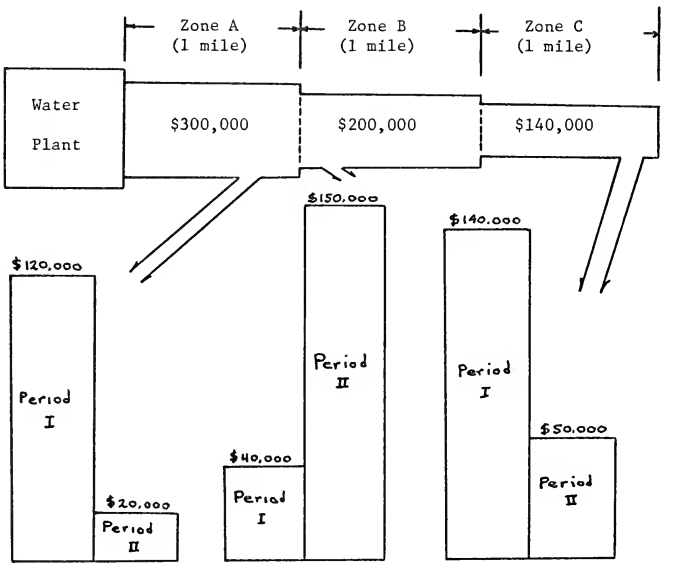
[23]Laurence C. Rosenberg, "Natural-Gas-Pipeline Rate Regulations: Marginal Cost Pricing and the Zone Allocation Problem," Journal of Political Economy, LXXV (April, 1967), pp. 159-68. For other relevant material see: Edgar M. Hoover, The Location of Economic Activity (New York: McGraw-Hill Book Co., Inc., 1948), pp. 1-187; August Losch, The Economics of Location, trans. by William H. Wogham (New Haven: Yale University Press, 1954), pp. 101-507.

[24]The formulation of the model rests entirely with Laurence C. Rosenberg and is not original in any part to the writer. The writer only adapts the model to the water industry and any errors in the model's application and interpretation are the writer's.

reflect the use characteristics of each of the zones. There is a peak period (Period I) and an off-peak period (Period II). The costs of the

Figure XXXII

Capacity Costs and Use Characteristics
of Three Zones



distribution system (water mains and storage tanks) are shown above. To serve the three zones, the utility must incur long-run marginal costs of $300,000 per mile in Zone A, $200,000 per mile in Zone B, and $140,000 per mile in Zone C. The capacity cost responsibility of each zone is shown in Table 1 below.

Table 1

Zone Share of Capacity Costs
(In 1,000's of Dollars)

Zone	A	B	C	Total
Zone A				
Period I	120 X 1	---	---	120
Period II	---	---	---	---
Zone B				
Period I	40 X 1	---	---	40
Period II	---	150 X 1	---	150
Zone C				
Period I	140 X 1	---	140 X 1	280
Period II	---	50 X 1	---	50
Total				
Period I	300	---	140	440
Period II	---	200	---	200

The capacity required in Zone A is $300,000 per mile to meet the system peak in Period I. The capacity required in Zone B is $200,000 per mile since this amount is the peak demand of Zones B and C occurring during Period II. The total capacity requirement of $300,000 in Zone A is the summation of each individual zone's requirement during the system's peak in Period I. The Zone B requirement of $200,000 in capacity represents the summation of the Period II demands in Zones B and C, which is their individual peak, and the system's peak in these two zones. During Period I, the capacity of Zone B need be only $180,000. Because of the character of the various demands, the capacity costs in Zone B have to be

divided between Zones B and C customers on the basis of their consumption during Period II.[25] The $140,000 of capacity in Zone C is needed by Zone C customers only, and, consequently, this amount is charged entirely to them during Period I. Using this method of cost allocation, each user pays in accordance with the costs he imposes upon the system. The allocation is determined by the proportion of the water taken during the peaks in each of the zones in which they are involved. The total capacity in Zone B is that needed to serve both Zones B and C. The excess capacity in Zones A and B with respect to the water used in Zones B and C are irrelevant externalities to each of the respective zones. The capacity in these two zones is greater than what is needed by the customers in these two zones, but it is required to serve Zone C.

Using the model in Figure XXII, the cost allocation can be altered if the users in each of the zones change their consumption patterns once the distribution system is installed. An increase in consumption during Period II by Zone C customers has no effect on the users in the other zones. The only effect is a reduction in water pressure to all the users in Zone C. As long as these users are willing to tolerate the reduced pressure, no change in rates is required. However, if users in Zone B increase their use during Period II, they can get additional water equal to the water flowing into Zone C. Zone C users will experience pressure losses until

---

[25] Rosenberg, "Natural-Gas-Pipeline Rate Regulations: Marginal Cost Pricing and the Zone Allocation Problem," p. 164.

there is no longer a water flow into the zone. The external costs of an increase in Zone B's users' consumption requires rates in Zone B to be increased during Period II so as to reduce their consumption level to the $150,000 of available capacity. If users in Zone A increase their use in Period II, there would be no call for an increase in Zone A rates since $80,000 of excess capacity exists in that zone during this period. The $80,000 of excess capacity in Zone A is also the system's excess capacity during Period II. Therefore, the use of the excess capacity in Zone A by Zone B customers has no effect on the water service in Zones B and C. However, if customers in Zone A increase their use of water during Period I, the water service to Zones B and C will deteriorate. There is an external cost associated with the increase in Period I consumption by Zone A customers since Zones B and C require $180,000 of capacity. If Zone B customers increase their use during Period I, the service costs associated with the increase in consumption by Zone B users requires their rates to be increased. If the Zone C customers increase their water use during Period I, there is no effect on any other zone. Therefore, no adjustment in rates is required.

Under the arrangement described above, each user (zone) contributes to the capacity costs in accord with the water used during the peaks of the various zones. In this example, all zones contribute to the capacity costs in Zone A during the system peak. Zones B and C contribute to the capacity costs in Zone B in accordance with their peak use occurring in Period II. Zone C is the only zone assessed for the capacity costs in Zone C.

A zone pricing solution can be used for rate differentials in both seasonal and daily peak load problems. The capacity costs during the seasonal peaks can be built into the block-rate differentials in each of the zones for each month of the year in accordance with the use during the seasonal peak. The rate differentials can be used to reflect hourly peak and off-peak costs and the zones' relationship to the system hourly peak and off-peak demands. In the case of adjusting costs by zones, the sale of water to users in Zone B by users in Zone A, or the sale of water to users in Zone C by users in Zone B, can be prevented. The municipal water company is given a monopoly to produce water for all three zones. A user in Zone B, selling water to a user in Zone C (assuming the rates are lower in Zone B), makes himself a water supplier in competition with the municipal water utility. This activity is illegal under the terms of the franchise. People in different zones, but adjacent to each other, can also be checked through their water consumption. A user in Zone B supplying water to a user in Zone C would have an excessively high amount of water use, and the user in Zone C would have an excessively low amount of water use. Also, if a user in one zone attempts to take enough water to serve himself plus a neighbor in the next zone, a larger meter is required so the two can receive the same quality of service as they received when both paid for their service separately. The use of a larger meter requires a greater service charge to be paid by the user supplying the water since the ability to congest the system has been increased.

Zone pricing can also be used to adjust rates in accordance with cost differentials arising from factors such as geographical characteristics and

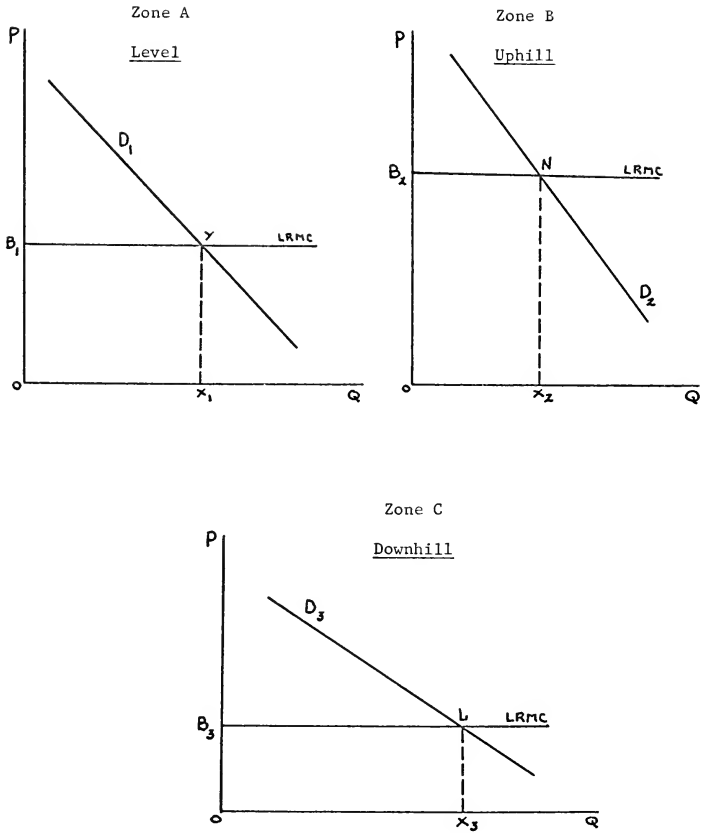
population density. The application is the same as the allocation of the capacity costs. These costs are assignable to each of the zones since a joint cost problem does not exist as is present in the peak and off-peak demand case.

If the utility starts operations by serving Zone A, a given amount of capacity is required. As expansion takes place into Zone B, which is all uphill, additional capacity will be needed. The new capacity will consist of pumping stations, additional mains, and storage facilities in Zone B plus greater capacity through Zone A. However, the additional capacity in Zone A does not increase the supply of water to the users in Zone A because, if the users in Zone A take more water, the quality of service available to Zone B customers declines. As mentioned earlier, this deterioration in quality is a social cost associated with Zone A users' increase in consumption, and, for efficiency purposes, requires a rate increase to Zone A customers. The fact that costs are assignable to each zone is demonstrated in Figure XXXIII below where the long-run marginal costs of serving the three zones are presented. The long-run marginal costs of each zone include additional operating costs and the additional capacity costs--the additional capacity in the zone itself plus the additional capacity which must be added to the other zones. The amount of capacity and the price are determined by the intersection of the long-run marginal cost function and the demand function.

If a peak and an off-peak demand are present in each of the zones, the pricing solution requires each period's price to be equated with the short-run marginal cost function (not shown in Figure XXXIII) and the demand

Figure XXXIII

The Marginal Solution for
Serving Three Zones



curves can be interpreted as the average demand or the weighted sum of the peak and off-peak demands. Under the assumption of constant costs, the sum of the prices is equal to long-run marginal cost and the demand functions, as in the Williamson solution, represent the weighted sum. The price OB_1 in Zone A, OB_2 in Zone B, and OB_3 in Zone C are the average prices that will be required of each zone's customers over the entire time horizon if a Pareto optimal solution is to be reached. The total costs and total revenues of serving each of the zones are OB_1YX_1 in Zone A, OB_2NX_2 in Zone B, and OB_3LX_3 in Zone C. The failure of utilities to use zone pricing in a manner described in Figure XXXIII results in a system of inter-area subsidies. The usual solution for determining a single price would be to take a weighted average of the costs in each zone and use this average as the single price. This alternative would lead to over expansion in some zones where the price is less than marginal cost, and under expansion in the zones where price is greater than marginal cost. The amount of capacity in any zone will not be Pareto optimal unless some zone has costs equal to the average price.[26]

---

[26]The zones in Figure XXXIII are assumed to start at the site of the water plant and are served in succession. Zone A will be served before Zone B and Zone B before Zone C. However, if Zone C were to be served before Zone B, a joint supply problem seems to appear but such is not the case. To get service into Zone C, the utility must pass through Zone B, thus providing capacity to Zone B. However, if Zone B users were to take water, Zone C users would not receive water due to pressure losses. Greater capacity is needed through Zones A and B to serve both Zones B and C. The most likely solution is that the utility, upon passing through Zone B, will provide enough capacity in that zone so that it can be served in the future. Until service is actually provided in Zone B, Zones A and C will have to carry the costs of the excess capacity in Zone B. However, the rates charged in Zone B can be set to provide a compensating subsidy

Conclusions

Through the use of economic analysis and the tools of welfare economics, it has been possible to develop theoretical models showing alternative solutions under various distributions of surplus. It was demonstrated that a single price monopoly solution results in the smallest size plant which can be rationalized on economic criteria. The use of perfect first degree price discrimination results in the same output and total surplus as a competitive solution, but the distribution of the surplus must be judged on non-economic criteria. The economics of block pricing were examined and it was demonstrated that the number of blocks determines the proximity of actual capacity to the welfare maximizing capacity. However, it appeared that these models do not fully explain the techniques employed by municipal water utilities. Consequently, a model was developed to show the alternative of quantity maximization combined with the out-of-pocket cost pricing as having greater relevance. The welfare and equity implications of these alternative models were examined, and it was concluded that the welfare effects were some of the worst.

To increase the general applicability of the theoretical models, and to make the rate structures conform more closely to economic efficiency

---

to Zones A and C so that, in the long-run, all zones pay a price equal to the additional costs of their service. If Zone B takes water without additional capacity being added, the prices in Zones B and C will have to be increased. The rate increases, mutatis mutandis, will be equal to the social costs of their congestion which will be equal to the capacity costs of providing each zone with the proper amount of capacity to meet their demands where price is equal to long-run marginal cost.

criteria, some adjustments in rate practices were suggested. The use of a "net" demand curve provided the means of establishing a fixed charge to overcome the problem of a shifting peak when hourly demands are relevant for pricing, while at the same time approaching a solution which is welfare maximizing in the Williamson sense. The fixed charge, based upon the meter size which measures the ability to congest the water system, provides the means of recovering the capacity costs which cannot be controlled through the use of prices because of the interdependency of demands, multiple price changes, and metering costs. The seasonal demands can be controlled through a variation in the variable charge as suggested by the economic models.

The use of zone pricing was recommended as a means of making rate structures conform more closely to efficiency criteria. The use of zone pricing enables rate structures to more accurately reflect marginal costs through their demand characteristics, geographical characteristics, and population density. The use of zone pricing better enables the attainment of the price equal to marginal cost criterion. The conclusion was also reached that, contrary to the conclusions of several economic models, in some cases, users taking water during the system's off-peak should be required to pay part of the capacity costs.

However, in the case of municipally owned water utilities, distributive judgments are built into rate structures when utilities are placed in the entire framework of the municipality's revenue-expenditure budget. The problem becomes one of showing the efficiency implications of these distributive judgments and the means of financing deficits created by

quantity maximization policies. These are the objectives underlying the following chapter.

CHAPTER VI

DISTRIBUTIVE JUDGMENTS AND ECONOMIC EFFICIENCY

Introduction

It is obvious that municipal water utility officials do not adhere rigorously to economic efficiency criteria in the determination of utility rate structures. Rate policies do not (and perhaps should not) attempt to maximize welfare in the Paretian sense. Rate structures generally incorporate block pricing usually justified on a cost basis, but the cost criterion is primarily average or fully distributed costs.[1] Municipal water utilities make little or no use of rate differentials which reflect the different cost impacts of the peak and off-peak demands. However, it cannot be said that municipal water rate practices are founded on completely irrational criteria. There is a rationale for the rate practices that are pursued regardless of how "irrational" these criteria may appear to the theorist.

Municipal water revenue is one of several sources of revenue available to communities for financing municipal services. It is inevitable, therefore, that water rate structures will have a relationship to local financial and development policies. The concern over social objectives such as tax efficiency, greater revenue, and community growth and

---

[1] For an excellent statewide survey of municipal water utility practices see: Staff Report, A Study of Municipal Water and Sewer Utility Rates and Practices in Georgia (Atlanta: Georgia Municipal Association, 1965).

development requires distributive judgments about the financing of a municipal water utility and leads to rate practices that might deviate considerably from the criteria that have been developed in earlier chapters.

The purpose of this chapter is to demonstrate how distributive judgments might cause utility practices to depart from efficiency criteria, and to demonstrate how economic analysis might provide some basic guidelines to preserve efficiency even though other objectives are carried out through the use of water rate structures.

Water Rates and Tax Policy

Many local communities frequently are forced to rely upon their water utility as a source of revenue for financing general fund expenditures because of state preemption of income, sales, and other taxes, because of constitutional restrictions and other limitations on taxing powers, and because of local notions of tax equity. When a water utility is used for revenue purposes, water rates may incorporate a tax component. Therefore, the implications of the "tax" must be evaluated in the light of tax equity.

The two major alternative criteria for the attainment of tax equity are that each individual be taxed (a) in accordance with ability to pay, or (b) in accordance with benefits received. To be equitable, a tax should be levied in such a manner that those individuals with equal abilities to pay, or equal benefits received, pay equal taxes. As a corrollary, those individuals with unequal abilities to pay, or receiving

unequal benefits, should be taxed unequally.[2] Attempts to attain tax equity have led to some municipal water rate practices heretofore labeled as arbitrary or "non-efficient."

Tax Efficiency

When a municipality relies upon its water utility as a source of revenue, the water rate structure is not constructed to raise a maximum amount of net revenue. The usual practice is to use the utility to raise an "expected" or budgeted amount of general revenue. It becomes significant to develop criteria that might be employed when the utility is obliged to provide a lump-sum amount of dollars to the general revenue fund of the community. The question is how to raise a given amount of dollars with the least loss in consumer surplus. Assuming that the amount of revenue the utility is budgeted to raise is equal to P_3EFB in Figure XXXIV, the utility may employ one of four alternatives: (1) rates may be set so that the utility's operations yield net profits equal to P_3EFB; (2) an ad valorem tax with a yield of P_3EFB might be placed on the water bills; (3) a per-unit water tax yielding P_3EFB might be levied; or (4) a fixed service charge yielding P_3EFB might be employed.

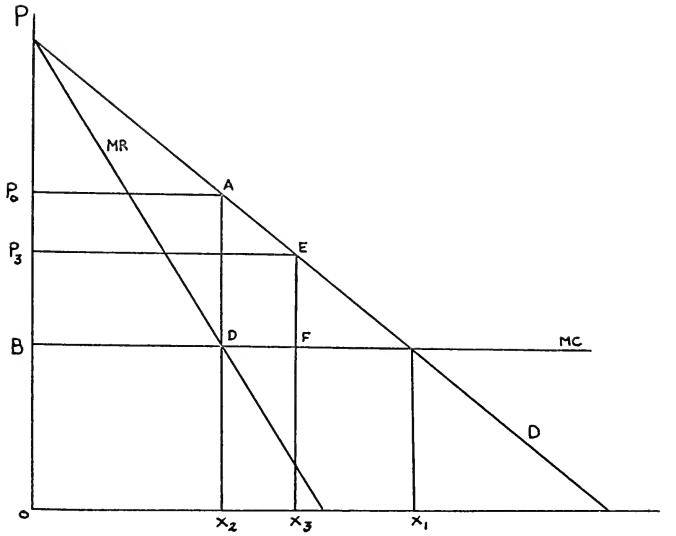
The first alternative, which entails operating with positive but limited profits, is, in effect, the same as setting a profit minimum and determining the rates which will provide this minimum. The use of a model

---

[2]For development of these points see: James M. Buchanan, The Public Finances (Homewood: Richard D. Irwin, Inc., 1960), pp. 165-75; John F. Due, Government Finance (3rd ed.; Homewood: Richard D. Irwin, Inc., 1963), pp. 102-21.

Figure XXXIV

A Municipal Water Utility with
a Profit Restraint



for this alternative is demonstrated in Figure XXXIV where D is the demand for water and MC is the marginal cost of supplying water.[3] The optimal output is OX_1 and the simple non-discriminating monopoly output is equal to OX_2. However, the firm is operating under a profit requirement established by local authorities and the simple monopoly output yields profits

[3]This model is an adaptation of a model developed by Baumol. See: William J. Baumol, Business Behavior, Value and Growth (rev. ed.; New York: Harcourt, Brace and World, Inc., 1967), pp. 1-79.

(P_oADB) which exceed the profit requirement of P_3EFB. In order to keep profits equal to the constraint, the utility expands output to OX_3 and reduces price to OP_3. Output X_2X_3, the output in excess of the simple monopoly profit-maximizing output, yields a marginal revenue which is less than marginal cost, thus reducing profits to the required amount (P_3EFB). The per-unit profit (the tax received by the city) is equal to EF. Since the total profit can be converted into a per-unit profit, the effect of this alternative is the same as a per-unit tax with a yield of P_3EFB, and the basic question becomes one of determining the comparative effects of a per-unit tax, an ad valorem tax, and a fixed service charge (a lump-sum tax) designed to provide equal yields. The basic models for this type of comparison have been developed by Richard A. Musgrave.[4] The loss of consumers' surplus can be measured by the increase in price which results from the tax. Consequently, economic efficiency would require that tax to be levied which obtains the same yield with a lesser increase in price and a smaller reduction in output from the optimum output.[5] The relevant case under consideration is a monopoly firm which, under municipal control, would produce the output where price equals marginal cost. In Figure XXXV below, NR is the demand for water and MC is the marginal cost of supplying water. If the utility is not required

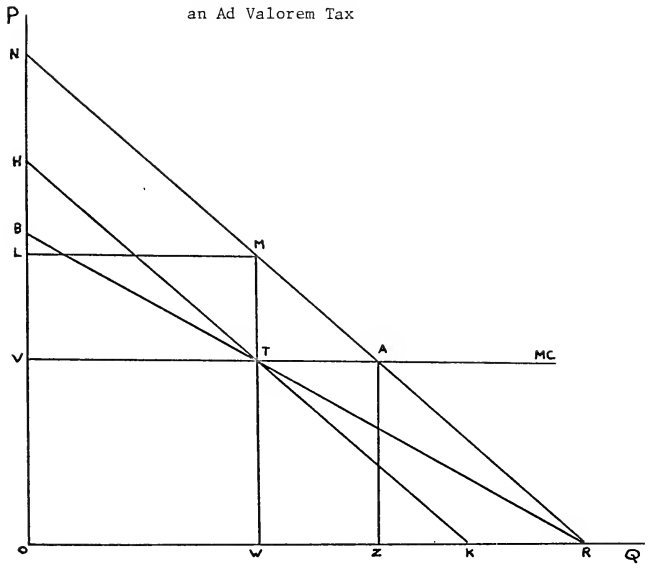
---

[4]Richard A. Musgrave, The Theory of Public Finance (New York: McGraw-Hill Book Co., 1959), pp. 276-311.

[5]Ibid., p. 302.

Figure XXXV

A Per-Unit Tax versus
an Ad Valorem Tax



to provide general fund revenue, the optimal price-output relationship is OV price and OZ output. However, when a revenue requirement is imposed upon the utility, the price-output relation changes. When a per-unit tax is levied, the amount of the tax (NH in Figure XXXV) can be shown by a decrease in the demand curve from NR to HK. Thus, with the imposition of a per-unit tax, output decreases to OW, price increases to OL, and the tax yield is equal to LMTV. An ad valorem tax of equal yield will shift the

demand curve from NR to BR. If the tax yield is to equal that yield under a per-unit tax, the demand curve must pass through point T. Consequently, the price increases to OL, output decreases to OW, and the tax yield is equal to LMTV. Therefore, an ad valorem tax or a per-unit tax, levied on the output of a monopoly utility operated as though it were a competitive firm (that is, price is equal to marginal cost), have the same impact on price, output, and consumers' surplus. The third effective alternative, a tax built into the fixed service charge, can be argued to be the better alternative. Since the tax is effectively levied on the infra-marginal units as a lump-sum tax, on the assumption of zero income effects, the tax will have no effect on the margin. Consequently, the price-output relation of OV price and OZ output is maintained. Therefore, the use of the fixed service charge will extract the amount of revenue budgeted from the utility while at the same time maintaining the optimal price and output of water. The fixed charge becomes the most efficient form of taxation and the one which should be given serious consideration by water rate analysts.[6]

In an earlier part of the present chapter it was stated that the alternatives of a per-unit tax and a profit restraint were identical in their

---

[6]These conclusions obviously are derived from an analysis based on an assumption of constant long-run costs. However, it should be noted that under an assumption of either increasing or decreasing costs, the equivalence between marginal and average costs is not maintained, and profits or losses come into existence under a marginal cost pricing policy. Consequently, the conclusions will be altered. In the presence of a profit restraint, where "natural" profit is not equal to the profit constraint, the analysis becomes entangled in a series of complexities involving a "trade-off" between taxes and profits.

effects. However, this conclusion is true only under the assumption of a simple non-discriminating monopoly. If block pricing is used, a difference in the impacts of the two alternatives develops. In referring back to Figure XXXV, the price and output effects of using block rates to raise a tax yield equal to LMTV will be defined by some point between point T and point A (this general effect was developed in Chapter V as an alternative between a simple monopoly and a perfect first degree discriminating monopoly). The exact location depends upon the characteristics of the block-rate structure. Consequently, block pricing will provide an output which is greater than the output which results from a per-unit tax and an ad valorem tax, but the output will be less than the optimal output maintained by the use of a fixed service charge tax. It appears, however, that block pricing is the most common form of charge water utilities employ.

When a community has very limited financial resources, it tends to exert its monopoly position in the sale of water and block pricing was demonstrated to be the most financially feasible alternative of putting this monopoly power into effect. Therefore, it is necessary to examine the tax significance of block pricing.

Block Pricing and Taxation

General revenue (profit) was maximized when the marginal revenue of the last block was equal to long-run marginal cost and output was sold in blocks of equal size (assuming a straight line demand function). The amount of general revenue the community wants to derive from its

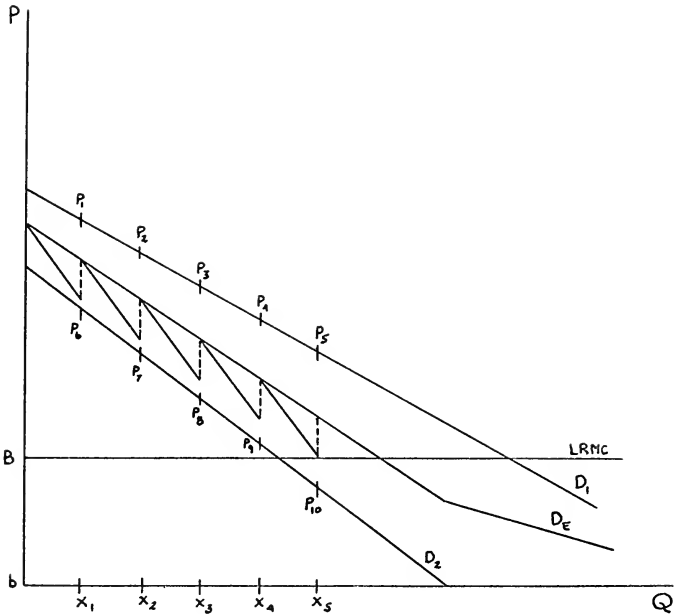
water utility determines the number of blocks put into the rate structure--the greater the desired level of revenue, given the demand functions, the greater the number of blocks. A model involving five blocks is shown in Figure XXXVI. Starting at point G where the effective demand intersects the long-run marginal cost function, the horizontal distance BG is divided into six equal parts. At output bX_5, effective marginal revenue is equal to long-run marginal costs. Output bX_1, the quantity in the first block, is sold at a peak price of X_1P_1 and an off-peak price of X_1P_6. The second block quantity (X_1X_2) carries a peak price of X_2P_2 and an off-peak price of X_2P_7. The third, fourth, and fifth blocks, X_2X_3, X_3X_4, and X_4X_5 respectively, are sold at X_3P_3, X_4P_4, and X_5P_5, the respective peak prices, and X_3P_8, X_4P_9, and X_5P_{10}, the respective off-peak prices. Output bX_5, the total output, determines plant capacity.

It is not necessary that blocks in the rate schedule be of equal size. The size and number of blocks might be varied in accordance with the tax objectives of the community. A rate structure consisting of a few blocks at the beginning of the schedule and numerous blocks toward the end of the schedule places a greater share of the revenue burden upon large users relative to small users. In other words, this type of schedule puts the greatest burden of municipal services on large commercial users, industrial users, and residential users doing extensive lawn sprinkling.

An allocation of the burden falling heavily upon large users might be used to place a large share of the "water-rate tax" upon individuals outside the community. When large users are primarily large industrial users, these firms probably sell a very small proportion of their output

Figure XXXVI

A Five-Block-Rate Schedule



to customers within the city. Most of their sales will be in a multi-state region outside the boundries of the community. Therefore, in these instances where the large users bear the greatest relative revenue burden, the tax might be passed on to their customers via higher prices and the burden is not shifted to people inside the community. By avoiding the intra-community tax shifting problem, the municipality avoids the distributional effect of some users paying, through higher prices, more

than their allocated share of the burden of the "water-rate tax."

A small number of blocks at the front of the schedule and few blocks at the end, or an early open-ended block, result in small users paying a relatively greater share of the revenue burden of the tax. This effect results from large users, having a greater proportion of their water use falling in the last block than small users, pay the lowest marginal and per-unit price. When this type of block structure is used, the tax burden will be the most regressive. However, it should be noted that a block schedule with a large number of blocks at the beginning of the schedule places the greatest relative burden on the largest number of users the utility serves and it is these users (residential) who impose the greatest extra-capacity costs upon the utility. Therefore, this type of rate schedule may be valid on economic grounds. However, the case currently under consideration is the use of rate structures as taxing devices. The rate structure with a large number of blocks at the front of the schedule imposes the greatest relative burden on residential users whose water use is primarily for domestic uses such as washing and laundering.[7] These uses have the lowest demand elasticity and a rate structure falling heavily upon these demands might raise two problems. The rates might distribute the tax burden in a manner which conflicts with the ability to pay. How-

---

[7]This type of schedule can be justified on economic grounds and therefore the extra-capacity costs, or demand charges, should be included in the first few blocks of the rate schedule. See: Albert P. Learned, "Financial Problems of Municipally Owned Water Utilities," Journal of the American Water Works Association, L (August, 1958), p. 1012. Also see: F.P. Linaweaver and John C. Geyer, "Use of Peak Demands in Determination of Residential Rates," Journal of the American Water Works Association, LVI (April, 1964), p. 409.

ever, if the community deems the most "equitable" basis for taxation is to charge what the traffic will bear, this rate structure is by definition the most equitable.[8] A second problem is setting rates on the first blocks so high that some very low income users might curtail their water use to levels below the amount considered necessary for minimum health standards. Although indoor plumbing is available, it may not be used--toilets need only be flushed once a day.[9]

When a community uses a single rate structure applied to all user classes (as suggested by efficiency criteria), all users pay the high per-unit charges in the early blocks of the rate schedule. However, revenue or tax objectives might justify different schedules for different user classes. A single rate schedule places a heavy burden on all users for their consumption of the first increments of the total quantity of water taken. The effect might be a distribution of the tax burden in conflict with tax objectives which determined the block sizes in the first

---

[8] Some writers feel that the criterion used by municipal water authorities for rate purposes is to charge what the traffic will bear. See: Jerome W. Milliman, "The New Price Policies for Municipal Water Service," Journal of the American Water Works Association, LVI (February, 1964), p. 127.

[9] The type of adjustments in the number of blocks herein described is very much in evidence in the block structure of the Athens, Georgia, rate schedule. The Athens schedule includes a disproportionate number of water blocks in the middle of the block schedule relative to other cities in Georgia. The explanation given to the writer by the city engineer is that Athens has a disproportionate number of middle sized users relative to other users than do other Georgia cities. No mention was given to the costs these users impose upon the system relative to other users. The basis for the blocks is obviously benefits. It was cited that the Athens water utility turns over about $1,000,000 annually to the city, and the middle sized users are expected to share in the burden of this revenue in proportion to their relative numbers.

place because no distinction is made between different user classes. This problem may be avoided by using a separate rate schedule for each class of user. Then the burden of the revenue tax can be placed directly on those users for whom it was intended (assuming there is no shifting). If the community wants the largest tax contribution from residential users, the average residential demand can be estimated and the blocks in the residential rate schedule adjusted accordingly. Therefore, the commercial and industrial users do not have to pay the high per-unit rates since they have their own rate schedule. However, if the burden is to be placed on large users, the blocks in the industrial and commercial user schedule can be altered with the residential user schedule being left alone. Therefore, the use of block schedules for each class of user enables the municipality to better carry out the tax objectives of the water rate schedule in a manner which is both more efficient and "equitable" than when a single rate schedule is applied to all classes. However, different rates to different classes involve the sacrifice of economic efficiency and is discriminatory pricing based upon "non-economic" criteria.

Problems Associated with the Use of a "Water-Rate Tax"

The use of a "water-rate tax" provides the only means available to many communities for raising revenues to finance municipal expenditures, or for extracting taxes from county residents to be used to finance municipal services from which the county water user cannot be excluded, such as purely collective goods. However, there are some serious drawbacks to this form of taxation which must be considered.

One problem associated with this form of taxation is the implicit assumption that the distribution of benefits, or the distribution of abilities to pay, is in proportion to the elasticity of demand for water. To the extent that this assumption is satisfied, the use of water bills as a form of taxation results in an equitable distribution of the tax burden. However, it appears doubtful that the distribution of benefits or abilities to pay conform to the demand elasticity of water use or to water bills calculated under a declining block-rate schedule. Therefore, a tax dependent upon elasticity, or the water bill, will not conform to tax equity criteria. A more equitable tax might very well be one which is built into the fixed service charge rather than taxes built into declining block rates, per-unit taxes, or ad valorem taxes.

Under a tax placed in the service charge, each water consumer makes a fixed contribution to the city's general revenues and the tax is not one based upon elasticity. The amount of tax included in the service charge might be based upon the proportion of other taxes each class of water user pays to the city, and thus achieve "horizontal" tax equity.[10] The use of this type of lump-sum tax will have little impact upon water use, an important consideration for purposes of minimizing so-called excess burden.

---

[10]The city might estimate what proportion of its tax revenues comes from residences, commercial businesses, and industrial users. These proportions can be applied to the total tax revenue the city wants to derive from the water users to estimate the share of the total tax to be allocated to residential user rates, the commercial user rates, and the industrial user rates. The tax itself can then be put into the fixed service charge part of the water bill.

A community is faced with a dilemma when using its water utility as a source of tax revenue. This dilemma arises between the use of water rates to achieve an efficient use of resources, and the use of water rates for tax purposes. A per-unit tax on the quantity of water used, or an ad valorem tax, which is directly related to water use, might have a significant impact upon water use because of the relative price elasticity of the demand for lawn sprinkling. Therefore, a per-unit tax can be used to bring about a reduction in water use for efficiency purposes by rationing water when water scarcities exist. However, for tax purposes, a reduction in water consumption represents an erosion of the tax base which provides the revenue needed for financing general fund expenditures. Therefore, careful administrative decisions are required to reach a compromise between the two alternative effects. On the other hand, the use of a fixed tax does not guarantee that each individual is being taxed in accordance with the benefits received, or the ability to pay. Each individual, contributing a fixed amount, might pay more or less than his benefits, or ability, depending upon the amount of use he makes of the collective good, or the size of his income. But the fixed charge eliminates the need for a choice between restricting water use to economize water supply and expanding water use to create a greater tax base.

Further Tax Efficiency Matters

Another form of taxation involving distributional implications is to have water rates which are not in line with costs. The failure of municipal water rates to incorporate zone pricing results in taxation, i.e., redistribution, although there may be no such intent. Water users living

in high geographical areas relative to the water plant usually pay the same rates as individuals living in relatively low areas. However, the cost of serving high areas is considerably greater (due to distributional problems) than the cost of serving low areas. When local users pay the same rates, part of the costs of serving high areas is being paid by users in low areas. Thus, rate relationships appropriate to a city located in a valley will be quite different from those of a city located on a plateau. Economic efficiency is being sacrificed,as the price of the marginal unit in high areas is below marginal production costs while the reverse is true in low areas. This system of implicit taxes and subsidies leads to the over expansion of service in high areas and under expansion of service in low areas which will have location implications within an area. Welfare could be increased by reducing prices in low areas and increasing prices in high areas. The specific distributional effects of this inefficiency are determined by the specific distribution of income and/or benefits between the two areas, depending upon which of the two tax bases are selected for equitable taxation.

If rate structures make no attempt to incorporate a charge based upon the time of water use, additional distributional effects are built into rate schedules. It has been demonstrated that costs of serving peak users are greater than costs of serving off-peak users. The former require additional capacity outlay while the latter make use only of existing idle capacity. If both groups of users pay the same rate, peak users are given a "tax"advantage in the sense that off-peak users are paying part of the costs of providing peak service.

Peak use is created largely by the water demands for lawn sprinkling and uncontrolled air conditioning during the summer months and these demands have relatively high income elasticity. Off-peak demand consists of occasional water use by residential users along with a large proportion of the water used by commercial and industrial users. The effect is to tax low income groups in order to serve peak users.[11] To achieve greater tax equity and to achieve greater economic efficiency, rate structures should reflect these differences in costs of service.

The Suburban Tax Problem

In many cases, the use of a "water-rate tax" provides the only means by which a community can reach suburban areas for tax purposes. "Inside" and "outside" water rate differentials usually are based upon the political boundries of the community. The rationale behind this jurisdictional distinction is frequently costs, but,more importantly, the justification is the attainment of tax equity and efficiency. It is generally recognized that those people living in the adjacent suburbs receive benefits from central city expenditures on such services as parks and recreation, street improvements, mosquito control, hospitals, and,in some cases, educational facilities such as libraries. However, these suburban beneficiaries are outside the taxing authority of the community. Thus, taxpayers residing inside the community may be subsidizing those taxpayers who live in the

---

[11] Most lawn sprinkling tends to occur in high income new suburban developments with large lots. The most stable loads are found in lower income older neighborhoods with small lots. These factors tend to indicate a situation where higher income groups receive a tax advantage over the lower income groups.

suburbs. The attempt to reach these suburban taxpayers can be carried out readily through the municipal water utility. The creation of water rate differentials between city and suburban residents in excess of the cost differentials of serving the two groups is one means of obtaining some degree of tax equity based upon benefits received.[12] The grounds for a benefit tax can be justified on economic criteria if the benefits received by non-taxpaying suburban residents from a municipally provided quasi-collective good are necessary for the determination of the Pareto optimal price and quantity. If these benefits are not considered, the supply of the public good will be under expanded.

Assume a community wishes to derive revenues from suburban users to finance a municipal service other than water. The supply of and demand for a quasi-collective good such as street maintenance provided by a central city government are shown in Figure XXXVII.[13] The horizontal axis measures the quantity of city provided goods, D_1 measures the "inside" demand for the collective good, and D_2 measures the "outside" demand for

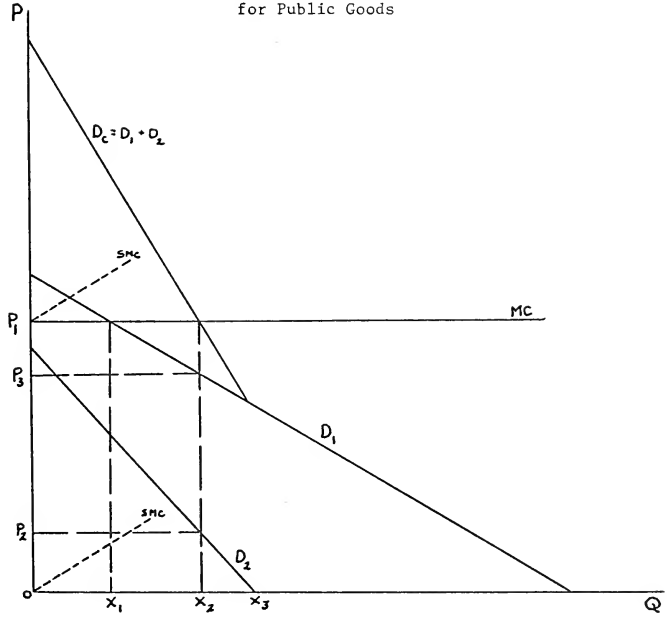
---

[12] In some cases, the suburban residents can be excluded from receiving benefits by exclusion from the facility. This exclusion is done in some communities by restricting the use of parks, golf course, and city dumps to the residents of the community. Admission is based on the place of residence. However, this exclusion cannot be carried out in the case of purely collective goods, such as street improvements, police protection, and mosquito control. The suburban residents cannot be excluded from the use of the service or the derivation of any benefits from the service.

[13] A purely collective good is one whose consumption by one individual does not diminish the total available to other individuals. See: Paul A. Samuelson, "The Pure Theory of Public Expenditure," Review of Economics and Statistics, XXXVI (November, 1954), pp. 387-89.

Figure XXXVII

Supply and Demand
for Public Goods



the good provided by the city.[14] At a price equal to marginal cost, inside city voters will vote for quantity OX_1.[15] However, since the good

---

[14]For a careful discussion of the demand and cost curves for a collective good, see: James M. Buchanan, Public Finance in Democratic Process (Chapel Hill: University of North Carolina Press, 1967), pp. 11-17.

[15]For this voting process, see: James M. Buchanan. Ibid., pp. 11-17, 144-68.

is available to county citizens, they will behave as though the good comes at a zero price. That is, quantity OX_3 will be "demanded" just by those county residents who are aware of a zero price. The Pareto optimal price relation is price OP_3 to city residents and price OP_2 to the county residents. The Pareto optimal quantity is OX_2. If quantity OX_1 is retained, congestion develops, service quality deteriorates, and social marginal cost (SMC) rises. Consequently, both city and county residents reduce their use (or consumption) of the facility until, at quantity OX_1, a voluntary adjustment is reached. But, in this equilibrium, the county residents have driven-off some of the city residents. Therefore, solution OX_1 is one where the city residents pay for the facility, both city and county residents use the facility, and there is congestion which results in either a reduction in the use of the facility or a deterioration in the quality of the service provided by the facility. That is, both classes get inferior service and the city residents pay for both. This result is in serious conflict with the Pareto optimal solution (OX_2) which requires both county and city residents to share the costs. Although congestion by both groups at quantity OX_2 may be difficult to avoid since both groups "behave" as though the use price is zero, equity is improved and facilities expanded if the county residents pay for part of the facilities.

Given an amount of dollars equal to the rectangle P_2BX_2O (in Figure XXXVII), which is the amount of tax revenue to be derived from the county residents, how can this revenue be collected with the least loss of consumers' surplus? The problem is basically one of institutionalizing the county payment. Using the four alternatives established in the previous

section, the payment can be extracted by: (1) setting outside rates so that utility operations are more profitable in the county than in the city; (2) using an ad valorem tax on the water bills of the county residents; (3) using a per-unit tax on the water consumed by county residents; or (4) using a fixed service charge tax. It was demonstrated that the most efficient alternative is the use of a fixed service charge tax since their is no change in the optimal output of water.

Community Growth and Development Policy

Industrial Location

Another primary concern of many communities is creating a local environment conducive to the economic growth of the community. The problems that have to be overcome are many, and, frequently, the solution to this problem is attempted through water rate practices. A factor determining the ability to attract industry is the quantity and quality of public services the municipality has to offer a potential industry and its employees. It is at this point that many small communities encounter difficulties. Smaller communities are faced with a deterioration in the property tax base, and further property tax increases may tend to abet further property deterioration. To encourage business location and reverse this tendency, these communities need to provide public service expenditures they cannot afford, given their present revenue sources. Communities in these circumstances frequently turn to their water utility as one of the means of encouraging industrial location. It is this attempt

which also explains some of the "irrational" rate practices which combine non-marginal rate patterns, in some instances, with marginal provisions, in other instances. For example, rate structures sometimes provide an off-peak discount designed to discourage peak use while at the same time provide low promotional rates which encourage peak use through uncontrolled air conditioning and other commercial and industrial uses.[16] The rate structure is constructed so that the last blocks, which are of concern to businesses, have prices below marginal costs, and, consequently, the utility provides industry with water service at below cost.

There appears to be an economic reationale for such low rates to industrial users. Many industries will be given other location incentives such as an exemption from personal property taxes and/or a partial exemption from real property taxes. Municipal water rates might be used by the community to recover these tax concessions through a readjustment of the blocks in the rate structure. In order to avoid this possibility, industries legitimately might require a community to give it preferential water rates with a guarantee that rates will not be increased.[17]

---

[16] William G. Shepherd, "Marginal Cost Pricing in American Utilities," Southern Economic Journal, XXIII (July, 1966), p. 64.

[17] A community which experienced the benefits from a favorable water schedule is Savannah, Georgia. They have not sold water to large users at rates below costs, but a separate utility was set up to supply only large industrial users on a contract basis. The costs of service are determined by dividing total costs by the total gallons pumped and each user then pays a per-unit price based upon this average cost of service. The residential and commercial water utility has kept its rates high and any community revenue needs have been pushed on to this utility's customers, enabling the large industrial users to avoid additional "taxation." The industries are also exempt from paying personal property taxes and only pay a nominal real property tax.

When tax concessions are given, someone has to bear the burden of the losses the utility incurs in the sale of water to these large users. The losses can be pushed on to the rates in the first blocks of the schedule so that water is sold at a price above marginal cost to many users. This alternative appears to be the one which is used to overcome losses incurred in the model developed in Chapter V (quantity maximization and out-of-pocket cost pricing). When such pricing practices are carried out, the utility tends to over expand, price of the marginal unit taken by small users is greater than marginal costs, and the price of the marginal unit taken by large users is below marginal costs. Efficiency can be improved by lowering prices and increasing the quantity available to small users, and raising prices and reducing quantity available to large users. The net effect is a reduction in capacity.

If the utility does not push losses from sales on to large users, the utility is operated at a loss and must receive a general fund subsidy. The additional revenue has to come from alternative sources such as property taxes and selective sales taxes. Both of these forms of taxation are regressive and the burden of the preferential treatment given to industry still falls hardest upon low income groups. However, if the losses are built into the early blocks of the water rate schedule, small users pay the highest per-unit price and the highest marginal price. This pricing scheme places the burden of the subsidy to large users on small users who appear to have the least elastic demands and the lowest average income. Therefore, the burden will be regressive with respect to income. But it seems doubtful that property taxes are as regressive as a water tax

built into the first few blocks.

Rate Differentials and Community Development

Through the use of zone pricing, rate structures can not only encourage industrial location in the manner discussed above, but they can also influence the direction of economic development within a community. Water rates, adjusted by various zones, can be used to achieve the long-range development plans of the community's long-range comprehensive plan. Areas in which development is to be encouraged can be given benefits from low water rates or special services. The type of development to be encouraged determines the class of rates to be given preferential treatment in each zone.

If the community has a given area in which they want to encourage residential development, low promotional rates to residential users can be offered while other users might be required to pay high penalty rates for locating in the same area. When water is offered to one class of user in one zone at below cost, the difference can be recovered through higher rates charged to other users in the same area, or the same class of users in other areas. In either case, users paying the higher rates are actually paying a type of social cost of locating in a zone in which their presence does not conform to the preferred use.

Because price is less than private marginal cost to users in zones in which development is being encouraged does not mean efficiency is being sacrificed. The relationship being satisfied is the sum (marginal private benefits + marginal social benefits = marginal social costs + marginal private costs). Although private benefits are less than marginal private

costs, the difference might be equal to the excess of marginal social benefits over marginal social costs. Going in the other direction, the users paying a price greater than marginal private cost may not be paying an unwarranted subsidy. There might be social costs connected with their location in certain zones. Although price is greater than private marginal cost, the difference is accounted for by an excess of social costs over social benefits, and, when users are paying a price equal to marginal private costs plus net social costs, welfare of individuals in the community is being maximized.

Water rates can also be used to encourage expansion of the center city and thereby prevent the development of urban slums. The provision of adequate zone pricing and preferential rates in areas around the center city enhances greater development and expansion of the center city. Growth is encouraged outward instead of upward. The development of urban slums might be prevented by discouraging commercial businesses from abandoning the center city and moving to adjacent satellite cities which leaves a vacuum conducive to urban decay.

One factor which must be considered is the burden of the development subsidy. When water rates are used for this purpose, users providing the subsidies needed to encourage growth are paying in proportion to their elasticity of demand for water unless the fixed service charge is used. Since it appears that water use increases less than proportionately to income increases, rate taxation has a regressive effect. If property taxes are used, the burden also is regressive. Although a comparison of the incidence of various types of taxes is not within the immediate scope of

this study, it is an important consideration to make when the administrative decision is made to use water rates as a means of encouraging area growth and development.

When water rates are used as the means of encouraging economic growth and development, their application must be tied closely to the planning programs of the city. Close coordination is necessary because, although development tends to follow the water supply, there are other variables which are as important, if not more important, in determining growth patterns. Some of these other factors are major lines of transportation, places of natural beauty such as lakes and parks, and open spaces. The improper use of water rate policy might lead to disorganization and community ill feeling about utility practices. The location of the water system might tend to isolate or deter growth in some areas. As in the case of Savannah, Georgia, the location of their industrial water plant adjacent to the existing industrial plants has deterred industrial location in other parts of the city where greater development has been desired. The utility was located on the north side of the city, and this location has prevented industrial development on the south side of the city, which, at the present, is an economically dead area.

The proper use of rate policy can have very beneficial effects on community growth and development. However, its improper application can have serious detrimental effects. Competent administrative judgment is required in its application.

Conclusions

Municipal water utilities provide a means of achieving community objectives such as tax equity and efficiency, and community growth and development. The use of water rates for these purposes leads to municipal water rate practices which sometimes appear to be "irrational"--at least in terms of strict economic criteria. Relevant economic models present solutions which equate marginal cost with price. The criteria applied for economic efficiency and welfare maximization are those that are applicable to privately owned utilities. However, the municipally owned water utility serves purposes other than supplying water in the most efficient manner. Distributive judgments are built into water rates as local government officials attempt to achieve public oriented goals. The purpose of this chapter has been to examine some of these social objectives that are sometimes built into water rates, how these objectives can be carried out, and their impacts upon the rate structure and income distribution. It was shown that there are four alternatives that can be used in raising revenue through a water utility. The utility can be operated in such a manner as to earn a budgeted "profit," an ad valorem tax can be applied to water bills, a per-unit tax can be applied to the water used, or a tax can be built into the fixed service charge. Of these four alternatives, it was demonstrated that in the case of a municipally owned utility being operated as though it were a competitive firm, the ad valorem tax and the per-unit tax have equal effects upon optimal price and output. The use of block pricing leads to deviations which are less from the optimum than the per-unit tax and the ad valorem tax. The use of a tax built into

the fixed service charge is the most efficient in that it does not entail a deviation of price and output away from the optimum, but yet is able to extract the necessary revenue. These conclusions were based on an assumption of constant costs.

It was shown that the construction of a block-rate structure, the usual form of water pricing, has significant distributional implications. Tax purposes may require different rate schedules for different users, whereas economic criteria requires all users to pay the same price. The use of the water utility also provides a means by which a city government may reach the county residents for tax revenues to be used in providing facilities from which the county residents receive benefits.

Implications in the use of water rates as taxing devices should be noted. The use of water bills as the basis for a tax effectively taxes individuals in proportion to their elasticities of demand for water. Large users may pay a relatively smaller share of the burden than small users because of the nature of block-rate structures.

The use of water rates as a means of carrying out programs of community growth and development were examined. The effects of water rates on the type and direction of industrial location and community development were examined.

When social objectives and their inherent distributive judgments are examined within the utility's role in the overall objectives of the community, practices which at first appear to be irrational do appear to have some social objective, regardless of how vain these rationales may be. This

chapter has presented, to a limited extent, the economic and welfare consequences of the various distributive judgments in the rate structure, an analysis seemingly entirely absent in the literature of economics and utility management.

CHAPTER VII

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Practitioners in the field of utility rate-making advocate both an "efficient" and an "equitable" solution to the pricing problem faced by water utilities. These criteria of efficiency and equity require the utility to cover its total costs and each user to pay a price equal to the so-called "full" costs he imposes. It becomes obvious that solutions based upon these criteria are cost-of-service oriented and require a determination of fully distributed average costs.

The usual technique of allocating costs on the basis of ratios between peak and average demands fails to satisfy the criteria of efficiency and equity established by practitioners as well as economists. Water rates are determined upon the basis of average demand and average costs, whereas costs vary according to the rate and quantity of output. As demand characteristics change, the ratio between peak and average demand changes and this, in turn, alters the cost allocation. Thus, an alteration in demand characteristics can affect the cost responsibility between peak and off-peak users without a change in capacity requirements taking place. The use of average demand fails to allocate to each user his "fair share" of costs because the technique itself creates inter-class subsidies. Also, by using averages and ratios, utility models fail to provide adequate solutions to the joint cost problem.

Even if joint costs are assumed away, utility methodology still must confront a capacity allocation problem. It appears that the basic problem with these models is their treatment of common costs as joint costs, i.e.,failure to realize the implications of vertical summation.

The use of average cost pricing leads to price structures which do not conform to the marginal conditions necessary for welfare maximization, i.e., price equal to marginal cost. Moreover, the failure to use peak and off-peak rate differentials and zone pricing further prevents the utility models from conforming to economic efficiency and creates redistributive effects between different groups of water users. These redistributive effects are not recognized for what they are: income transfers from one person to another.

Models based entirely upon economic efficiency criteria have advocated a modified form of marginal cost pricing, a conclusion derived through the geometric tool of the vertical summation of demand functions--Steiner, Williamson, Buchanan. Marginal cost pricing seems to provide an acceptable pricing guideline,although its use admittedly is limited by a number of underlying complicating factors and assumptions.

With the exception of Davidson (who advocates a price equated with long-run marginal cost), the theoretical models which deal with the peak load and joint cost problems described in this study develop the conclusion that proper pricing policy requires price to be equated with short-run marginal cost, whether this short-run marginal cost be money cost, opportunity cost, or social cost. However, this conclusion depends upon the satisfaction of some critical assumptions such as the

existence of independent demands, the absence of joint costs, and consistency between efficiency and distributional equity.

The original Hotelling model marked the introduction of several significant contributions to utility rate theory through the application of a marginal social cost criterion to the peak load problem. The Steiner model points to the necessarily discriminatory nature of prices in a joint cost solution, but he notes the Pareto optimality of the discrimination since the sum of the prices is equal to marginal cost. Hirshleifer attempts to redefine marginal cost in the presence of capacity constraints so as to eliminate the implication of price discrimination. This redefinition is accomplished through the use of marginal opportunity costs. These definitional questions became most obvious in the face of a capacity constraint where at least three different concepts of cost have been employed in defining the point of intersection between demand and short-run marginal cost. At outputs less than capacity ouput, the concept employed is money costs (out-of-pocket costs), and, at capacity output where price is equal to short-run marginal cost, the concept is either opportunity cost (Hirshleifer), social cost (Hotelling), or a rationing price (Williamson and Steiner).

When the assumption of independent demands is relaxed, the economic models encounter practical administrative problems which can be eliminated by the use of alternative pricing policies. The presence of demand interdependence is due to the fact that peak demand is a function of the off-peak price and the off-peak demand a function of the peak price. Consequently, when peak and off-peak rate differentials are used, the peak load might shift and the utility must readjust rates to keep price equal to

marginal cost. But, as prices are changed, the demand relationship between the peak and off-peak also changes, requiring still another price adjustment...and so on.

Models based on marginal cost pricing require the peak user to bear all capacity costs if the off-peak users fail to utilize capacity at an off-peak price equal to short-run marginal cost. This conclusion releases users who take no water during the system peak of any responsibility for capacity costs, thereby creating "free-riders" justified on efficiency grounds. This situation may be considered inequitable and, therefore, in the interest of "distributional equity," these users will probably be assigned part of the capacity costs. On the other hand, it can be argued, through an application of the Rosenberg model, that economic efficiency may require part of the capacity costs to be paid by the off-peak users, particularly in the case of distribution capacity.

Municipally owned water utilities are faced with a number of problems which have not been recognized (or at least not considered adequately) by models which attempt to provide solutions to the pricing and investment decisions faced by firms with a peak load problem. Some of these related problems are: a community's desire for growth and development; the pressure for increased general fund revenue; and the attainment of tax equity. In the attempt to carry out these social objectives, communities have employed water rate structures which have not always conformed to economic efficiency criteria. In attempts to use utility rate structures as a taxing device to raise revenue for the community,

or to reach heretofore untaxable suburban residents, communities have four alternatives: (1) to operate with a profit requirement; (2) to impose an ad valorem tax on water bills; (3) to impose a per-unit tax on the quantity of water used; or (4) to build a tax into the fixed service charge. The literature at both applied and theoretical levels fails to recognize the distributional implications of these possibilities.

Recommendations

It is obvious that water rates are not constructed in a manner designed to achieve welfare maximization, at least in the Paretian context. The apparent use of quantity maximization coupled with strong price discrimination leads to a vast over expansion in capacity and to a system of internal cross-subsidies that may aggrevate the distributional problem.

In the hope that water rate practices might conform more closely to economic efficiency (as well as becoming more equitable), several pricing and taxing guidelines may be recommended.

For a municipally owned water utility to properly evaluate the economic efficiency and distributional impacts of its rate practices, it is helpful to derive pricing guidelines based on efficiency criteria with distributional judgments put aside. This basic model can provide a standard of comparison for determining the distributional effects of rate practices based on efficiency criteria, the efficiency effects of rate practices based on distributional judgments, and the distributional

effects of equity judgments built into the rate structure. This type of analysis is extremely important in the case of municipally owned water utilities since, in their operations, economic efficiency and distributional equity are traded off as the two come into conflict. The model should be one which is based upon a set of assumptions which fit the conditions under which a water utility provides service. These assumptions are independent demands between seasonal peak and off-peak demands, interdependency between hourly peak and off-peak demands, and joint costs between both seasonal and hourly peak and off-peak demands.

In the case of independent demands, the Williamson model provides an optimal solution in terms of economic efficiency. His model entails equating effective demand with long-run marginal cost to determine plant capacity, and the off-peak price with short-run marginal cost (when the off-peak demand fails to utilize capacity fully at this price). The peak price is determined so that the quantity demanded is equated to the capacity to produce (assuming an indivisibility constraint).

The pricing scheme established by Williamson should be extended to take into consideration the costs of supplying various zones. This guideline can be used to determine the amount of capacity the utility should provide in each zone. The effective demand for capacity of each zone is equated with the long-run marginal costs of supplying each zone. These costs are composed of additional production costs, additional capacity costs in the zone itself, and additional capacity costs which must be spent in other zones as a result of additional service provided to the zone in question.

When demands are interdependent, administrative problems arise which can make the use of rate differentials difficult. However, the utility may, through the use of a fixed charge, overcome these problems and recover the cost differences in serving the hourly peak and off-peak users. The capacity costs can be estimated from engineering data which indicates the load imposed by various quantities of equivalent 5/8" meters. By using an average daily demand and the amount of water used during the peak hour, the additional capacity required to meet the hourly peak over and above the seasonal peak can be determined. The additional costs can be incorporated into a service charge which does not include water usage. The amount of the charge is determined on the basis of the number of equivalent 5/8" meters each user takes which bases the charge upon the ability of the user to congest the water system during the system's peak load and allows the user to select his own meter size and accompanying charge. Also, this alternative is recommended because of the present high costs of metering on an hourly basis, and it probably would more than off-set any gains which could be achieved through its implementation at the present time.

The fixed charge will not alter the pattern of water use, but it will prevent the problem of a shifting peak. There is no incentive for the users to shift their peak use over into hours which are, at the present, the off-peak hours. The individual has the freedom to select the combination of meter size and service charge he desires, and the utility has the ability to determine the degree of peak load capacity

it must provide. The service charge includes customer costs and extra-capacity costs involved in serving the hourly peak demand.

A variable use charge can be designed to recover the base water costs and the extra-capacity costs needed to serve the seasonal demand. The costs included in the variable charge should be spread evenly throughout the rate schedule with each water user paying the same marginal price. The total contribution of each user towards capacity costs is, therefore, proportional to the total quantity of water used. Consequently, block-rate structures, for efficiency purposes, should not be used. However, the use of block rates might be used when over-riding considerations entail the use of the utility as a revenue raising device, as a means of achieving some form of tax distribution, or as a means of promoting community development. But, the rate analyst should be aware of the many distributional effects of such rate structures.

The basic water charge should not include a water usage because the effect is to provide water to some water users on the basis of an effective flat monthly charge which entails elaborate bounties and cross-subsidies along with the inequality between price and marginal cost.

When water rates are used as a means of raising general fund revenues, the most efficient alternative form of taxation is the use of a tax built into the fixed service charge. This form of taxation is efficient in the sense that it does not result in a departure from optimal output, but, yet, enables the utility to extract the required amount of revenue. However, block pricing appears to be the most common form of pricing and it is probably fair to assume that any taxes will be

raised through this pricing mechanism. In this case, the taxing authority should be aware of the distributive implications of such taxation.

The use of separate rate schedules for different classes of users (although not recommended by strict economic efficiency criteria) might be used better to levy the tax burden on those users for whom it is intended. The use of separate schedules prevents all users from paying the high rates built into the block structure intended for a specific group of users.

The use of zone pricing and rate differentials might also be used to encourage industrial location and intra-community development which can enhance community growth and prevent slum development. In this instance, some sacrifice of pure economic efficiency must be made to encourage other social goals. However, social costs and benefits become an important consideration and they must be considered along with private costs and benefits to determine a social optimum.

An important consideration in the design of rate structures is the ability of the consumer to understand the basis for his water bill and the administrative costs of supplying such information. A water rate schedule, which is highly technical because of its adherence to strict economic efficiency pricing, can lead to user resistence and complaints about water bills. The effect is to increase the administrative costs of handling billing, as numerous bills must be explained and recomputed to satisfy customer complaints. Although overall efficiency might be increased by using a more simplified rate policy, the argument for simplicity is frequently employed to justify what in reality is an inequitable and inefficient rate structure.

BIBLIOGRAPHY

Articles

Averch, Harvey and Johnson, Leland L. "Behavior of the Firm under Regulatory Constraint," American Economic Review, LII (December, 1962), 1052-69.

Ayers, Louis E. "Methods of Financing Water Utilities in Michigan," Journal of the American Water Works Association, LI (January, 1951), 8-13.

Baxter, Samuel S. "Principles of Rate Making for Publically [sic] Owned Utilities," Journal of the American Water Works Association, LII (October, 1960), 1225-38.

Bergson, Abram. "A Reformulation of Certain Aspects of Welfare Economics," Quarterly Journal of Economics, LII (February, 1938), 310-34.

Boiteux, Marceo. "La Tarification des Demandes en Pointe: Application de la Théorie de la Vent au Coût Marginal," trans. H.W. Izzard, in Marginal Cost Pricing in Practice, ed. James R. Nelson. Englewood Cliffs: Prentice-Hall, 1964, 59-89.

Bonbright, J.C. "Fully Distributed Costs in Utility Rate Making," American Economic Review, LI (May, 1961), 305-12.

_____. "Major Controversies as to the Criteria of Reasonable Public Utility Rates," American Economic Review, XXX (May, 1940), 379-89.

_____. "Two Partly Conflicting Standards of Reasonable Utility Rates," American Economic Review, XLVIII (May, 1957), 386-93.

Bonine, E.D. "Making a Water Utility Solvent," Journal of the American Water Works Association, XLV (May, 1953), 457-58.

Boulding, Kenneth E. "Welfare Economics," in A Survey of Contemporary Economics I, ed. Bernard F. Haley. Homewood: Richard D. Irwin, Inc., 1952, 1-38.

Buchanan, James M. "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXX (August, 1966), 463-71.

Coase, R.H. "The Marginal Cost Controversy," Economica, XIII (August, 1946), 169-82.

Cornell, Holly. "Comprehensive Water System Plan," Journal of the American Water Works Association, LX (February, 1968), 125-28.

Crew, M.A. "Peak-Load Pricing and Optimal Capacity: Comment," American Economic Review, LVIII (March, 1968), 168-70.

Devel, Orville P. "Water Utility Rate Making," Journal of the American Water Works Association, LVIII (July, 1966), 845-48.

Faust, Raymond J. "The Needs of Water Utilities," Journal of the American Water Works Association, LI (June, 1959), 701-6.

Fisher, Gordon P. "New Look at Resources Policy," Journal of the American Water Works Association, LVII (March, 1965), 255-61.

Fouraker, Lawrence. "A Note on the Administration's Recent Tax Proposal," National Tax Journal, XVI (December, 1963), 426-28.

Fox, Irving K. and Herfindahl, Orric C. "Attainment of Efficiency in Satisfying Demands for Water Resources," American Economic Review, LIV (May, 1964), 198-206.

Gabor, Andre. "Further Comment," Quarterly Journal of Economics, LXXX (August, 1966), 472-80.

Grubs, Clifton M. "Theory of Spillover Cost Pricing," in Highway Research Record No. 47. Washington: Highway Research Board, 1964, 15-22.

Hatcher, Melvin P. "Basis for Rates," Journal of the American Water Works Association, LVII (March, 1965), 273-78.

Hicks, John R. "The Foundations of Welfare Economics," Economic Journal, XLIX (December, 1939), 696-712.

_____. "The Four Consumer Surpluses," Review of Economic Studies, XI (1943), 68-74.

_____. "The Generalized Theory of Consumer's Surplus," Review of Economic Studies, XIII (1945-46), 68-74.

_____. "The Rehabilitation of Consumers' Surplus," Review of Economic Studies, VIII (February, 1941), 108-16.

Hirshleifer, Jack. "Peak Loads and Efficient Pricing: Comment," Quarterly Journal of Economics, LXXII (August, 1958), 451-62.

_____ and Milliman, J.W. "Urban Water Supply: A Second Look," American Economic Review, LVII (May, 1967), 169-78.

Hopkinson, John. "On the Cost of Electric Supply," in The Development of Scientific Rates for Electric Supply. Detroit: The Edison Illuminating Co., 1915, 5-20.

Hotelling, Harold. "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," Econometrica, VI (July, 1938), 242-69.

Howson, Louis R. "Review of Ratemaking Theories," Journal of the American Water Works Association, LVIII (July, 1966), 849-55.

Kaldor, Nicholas. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," Economic Journal, XLIX (September, 1939), 549-52.

Keller, Charles W. "Design of Water Rates," Journal of the American Water Works Association, LVIII (March, 1966), 293-99.

Koplin, H. Thomas. "Discussion," American Economic Review, LI (May, 1961), 335-37.

Learned, Albert P. "Determination of Municipal Water Rates," Journal of the American Water Works Association, XLIX (February, 1957), 165-73.

_____. "Financial Problems of Municipally Owned Water Utilities," Journal of the American Water Works Association, L (August, 1958), 1009-13.

Linaweaver, F.P. and Geyer, John C. "Use of Peak Demands in Determination of Residential Rates," Journal of the American Water Works Association, LVI (April, 1964), 403-13.

Machlup, Fritz. "Theories of the Firm: Marginalist, Behavioral, Managerial," American Economic Review, LVII (March, 1967), 1-33.

Meyer, John R. and Kraft, Gerald. "The Evaluation of Statistical Costing Techniques as Applied in the Transportation Industry," American Economic Review, LI (May, 1961), 313-34.

Milliman, Jerome W. "The New Price Policies for Municipal Water Service," Journal of the American Water Works Association, LVI (February, 1964), 125-31.

Mishan, E.J. "On the Theory of Optimum Externality: Comment," American Economic Review, LVIII (June, 1968), 523-27.

Mohring, Herbert. "Relation Between Optimum Congestion Tolls and Present Highway User Charges," in Highway Research Record No. 47. Washington: Highway Research Board, 1964, 1-14.

Nelson, J.R. "Practical Applications of Marginal Cost Pricing in the Public Utility Field," American Economic Review, LIII (May, 1963), 474-81.

Patterson, William L. "Comparison of Elements Affecting Rates in Water and Other Utilities," Journal of the American Water Works Association, LVII (May, 1965), 554-60.

_____. "Practical Water Rate Determination," Journal of the American Water Works Association, LIV (August, 1962), 904-12.

Pauly, Mark V. "On the Theory of Optimum Externality: Comment," American Economic Review, LVIII (June, 1968), 528-29.

Pigou, A.C. "Railway Rates and Joint Costs," Quarterly Journal of Economics, XXVII (May, 1913), 535-36.

_____. "Railway Rates and Joint Costs," Quarterly Journal of Economics, XXVII (August, 1913), 687-92.

Rosenberg, Laurence C. "Natural-Gas-Pipeline Rate Regulations: Marginal Cost Pricing and the Zone Allocation Problem," Journal of Political Economy, LXXV (April, 1967), 159-68.

Ruggles, Nancy. "Recent Developments in the Theory of Marginal Cost Pricing, " Review of Economic Studies, XVII (1949-50), 107-26.

_____. "The Welfare Basis of the Marginal Cost Pricing Principle," Review of Economic Studies, XVII (1949-1950), 29-46.

Rynder, Arthur. "Demand Rates and Metering Equipment in Milwaukee," Journal of the American Water Works Association, LII (October, 1960), 1239-43.

Samuelson, Paul A. "The Pure Theory of Public Expenditure," Review of Economics and Statistics, XXXVI (November, 1954), 387-89.

_____. "Welfare Economics and International Trade," American Economic Review, XXVIII (June, 1938), 259-72.

Scitovsky, Tibor, "A Note on Welfare Propositions in Economics," Review of Economic Studies, IX (1941-1942), 77-88.

_____. "Two Types of Externalities," Journal of Political Economy, LXII (April, 1954), 143-51.

Shepherd, William G. "Marginal Cost Pricing in American Utilities," Southern Economic Journal, XXIII (July, 1966), 58-70.

Skelton, Garland G. "Factors Affecting Rate Increases and Revenue Bond Issues," Journal of the American Water Works Association, L (July, 1958), 919-22.

St. Clair, G.P. "Congestion Tolls-An Engineers Viewpoint," in Highway Research Record No. 47. Washington: Highway Research Board, 1964,

Steiner, Peter O. "Peak Loads and Efficient Pricing," Quarterly Journal of Economics, LXXI (November, 1957), 585-610.

Strotz, R.H. "Two Propositions Related to Public Goods," Review of Economics and Statistics, XL (November, 1958), 329-31.

Taussig, F.W. "Railway Rates and Joint Costs Once More," Quarterly Journal of Economics, XXVII (February, 1913), 378-85.

_____. (Untitled Rebuttal to A.C. Pigou), Quarterly Journal of Economics, XXVII (May, 1913), 536-38.

_____. (Untitled Rebuttal to A.C. Pigou), Quarterly Journal of Economics, XXVII (August, 1913), 692-93.

Troxel, Emory. "Incremental Cost Determination of Utility Rates," Journal of Land and Public Utility Economics, XVIII (1942), 458-67.

Vickery, William S. "Some Implications of Marginal Cost Pricing for Public Utilities," American Economic Review, XLV (May, 1955), 605-20.

_____. "Some Objections to Marginal Cost Pricing," Journal of Political Economy, LVI (June, 1948), 218-38.

Wallace, Donald H. "Joint Supply and Overhead Costs and Railway Rate Policy," Quarterly Journal of Economics, XLVIII (August, 1934), 583-619.

Walters, A.A. "The Theory and Measurement of Private and Social Costs of Highway Congestion," Econometrica, XXIX (October, 1961), 676-99.

Williamson, Oliver E. "Peak-Load Pricing and Optimal Capacity under Indivisibility Constraint," American Economic Review, LVI (September, 1966), 810-27.

Winch, David M. "Consumer's Surplus and the Compensation Principle," American Economic Review, LV (June, 1965), 395-423.

Wolff, J.B. "Forecasting Residential Requirements," Journal of the American Water Works Association, XLIX (March, 1957), 225-34.

Wright, Arthur. "Cost of Electricity Supply," in The Development of Scientific Rates for Electric Supply. Detroit: The Edison Illuminating Co., 1915, pp. 31-52.

Books

American Water Works Association. Water Works Practices. Baltimore: The Williams and Wilkens Co., 1925.

Arrow, Kenneth J. Social Choice and Individual Values. New York: John Wiley and Sons, 1951.

Baumol, William J. Business Behavior, Value and Growth. Rev. ed. New York: Harcourt, Brace and World, Inc., 1967.

_____. Economic Theory and Operations Analysis. 2nd ed. Englewood Cliffs: Prentice-Hall, 1965.

_____. Welfare Economics and the Theory of the State. 2nd ed. Cambridge: Harvard University Press, 1965.

Bergson, Abram. Essays in Normative Economics. Cambridge: Harvard University Press, 1966.

Bonavia, Michael R. The Economics of Transport. New York: Pitman Publishing Corp., 1936.

Boulding, Kenneth E. Economic Analysis: Microeconomics. New York: Harper and Row, Publishers, 1966.

Bowen, Howard R. Toward Social Economy. New York: Rinehart and Co., Inc., 1948.

Buchanan, James M. Public Finance in Democratic Process. Chapel Hill: University of North Carolina Press, 1967.

_____. The Public Finances. Homewood: Richard D. Irwin, Inc., 1960.

Clemens, Eli. Economics and Public Utilities. New York: Appleton-Century, 196

Davidson, Ralph K. Price Discrimination in Selling Gas and Electricity. Baltimore: Johns Hopkins University Press, 1955.

Due, John F. Government Finance. 3rd ed. Homewood: Richard D. Irwin, Inc., 1963.

Graaff, J. deV. Theoretical Welfare Economics. Cambridge: Cambridge University Press, 1963.

Hicks, John R. A Revision of Demand Theory. London: Oxford University Press, 1956.

_____. Value and Capital. 2nd ed. Oxford: Oxford University Press, 1946.

Hirshleifer, Jack, DeHaven, James C., and Milliman, J.W. Water Supply. Chicago: University of Chicago Press, 1960.

Hoover, Edgar M. The Location of Economic Activity. New York: McGraw-Hill Book Co., Inc., 1948.

Kafoglis, Milton Z. Welfare Economics and Subsidy Programs. Gainesville: University of Florida Press, 1961.

Lerner, Abba P. The Economics of Control. New York: The Macmillan Co., 1944.

Little, I.M.D. A Critique of Welfare Economics. Oxford: Oxford University Press, 1958.

Locklin, D. Philip. Economics of Transportation. 6th ed. Homewood: Richard D. Irwin, Inc., 1966.

Losch, August. The Economics of Location. Translated by William H. Wogham. New Haven: Yale University Press, 1954.

Marshall, Alfred. Principles of Economics. 8th ed. London: Macmillan and Co., Ltd., 1920.

Musgrave, Richard A. The Theory of Public Finance. New York: McGraw-Hill Book Co., 1959.

Pigou, A.C. The Economics of Welfare. 4th ed. London: The Macmillan and Co., Ltd., 1932.

Reder, A. Studies in the Theory of Welfare Economics. New York: Columbia University Press, 1947.

Samuelson, Paul A. Foundations of Economic Analysis. New York: Atheneum, 1965.

Smith, Adam. The Wealth of Nations. Modern Library. New York: Random House, Inc., 1937.

Stigler, George J. The Theory of Price. Rev. ed. New York: The Macmillan Co., 1962.

Vickery, William S. Microstatics. New York: Harcourt, Brace and World, Inc., 1964.

Williamson, Oliver E. The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm. Englewood Cliffs: Prentice-Hall, 1964.

Unpublished Material

Kafoglis, Milton Z. "Output of the Firm under an Earnings Restraint." Unpublished manuscript, University of Tennessee, 1968.

_____. "The Public Interest in Utility Rate Structures." Unpublished and undated manuscript.

Other Sources

Fox, Irving K. New Horizons in Water Resources Administration. Washington: Resources for the Future, Inc., 1965.

Howe, Charles E. and Linaweaver, F.P., Jr. The Impact of Price on Residential Water Demand and Its Relation to System Design and Price Structures. Washington: Resources for the Future, Inc., 1967.

Maryland. Guidelines for Improving Maryland's Fiscal Structure. Interim report, January, 1965.

Staff Report. A Study of Municipal Water and Sewer Utility Rates and Practices in Georgia. Atlanta: Georgia Municipal Association, 1965.

Staff Report. "The Water Utility Industry in the United States," Journal of the American Water Works Association, LVIII (July, 1966), 772-76.

BIOGRAPHICAL SKETCH

Robert L. Greene was born May 17, 1935, at Butler, Pennsylvania. In June, 1953, he was graduated from Butler Area Joint Senior High School. June, 1957, he received the degree of Bachelor of Arts with a major in Economics from Allegheny College. From 1957 to 1958 he attended the Pennsylvania State University, where he received the degree of Master of Arts with a major in Economics while serving as a graduate teaching assistant. During the period 1959 to 1963, Mr. Greene was an Instructor of Economics at Washington and Jefferson College. In 1963 he enrolled in the Graduate School of the University of Florida. He worked as an Interim Instructor in the Department of Economics until August, 1967. From September, 1967, until the present time he has been employed by the Department of Finance and the Institute of Government at the University of Georgia as an Assistant Professor.

Robert L. Greene is married to the former Heather Reid Hutchison and is the father of two children. He is a member of Pi Gamma Mu, Omicron Delta Epsilon, and the American Economics Association.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Business Administration and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1968

Dean, College of Business Administration

Dean, Graduate School

Supervisory Committee

Chairman